Identification of Linear and Nonlinear Systems using Multisine Signals

with a gas turbine application

---

by Ceri Evans

Prifysgol Morgannwg
The University of Glamorgan

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

April 1998

Abstract

Modern system identification techniques allow dynamic models to be directly estimated from measured data and the design of the data gathering experiment is a key step in any identification procedure. This thesis deals with the design of test signals for both linear and nonlinear modelling and their application to an engineering problem. It is motivated by a desire to fully exploit the recent advances in computer technology, which make the design and application of complex multisine test signals a practical possibility. The thesis can be divided into two parts, the first dealing with test signal design and the second presenting a detailed study of the testing and modelling of an aircraft gas turbine.

The main contributions of the first part deal with the influence of noise and nonlinearities on multisine test signals and the design of new types of multisines for testing nonlinear systems. The test times associated with single sine, multisine and maximum length binary signals are studied, with the aim of reducing test times while maintaining accuracy in the presence of noise. A novel methodology is presented for analysing the influence of system nonlinearities on multisines, with the aim of designing signals which are robust to nonlinear effects. This leads to the design of signals which can be used to identify the best linear approximation of block-oriented nonlinear systems of the Wiener-Hammerstein type. The design of signals which minimise the nonlinear distortion at the test frequencies is also studied, with the aim of identifying the underlying linear dynamics of the system. A scheme is proposed for the identification of linear systems in the presence of nonlinear distortions. The designs are then further developed to allow the direct measurement of points on the frequency-domain Volterra kernels (higher-order frequency response functions) of a nonlinear system.

The second part of the thesis deals with gas turbine modelling, with the aim of estimating models which can be used to verify the linearised thermodynamic models derived from the engine physics. The design of appropriate test signals is discussed, a detailed analysis of the measured data is presented and engine models are identified. The influence of noise and nonlinearities on the estimated models is studied. It is shown that the use of multisine signals and frequency-domain techniques is particularly suited to this problem, since the continuous-time s-domain models needed to validate the thermodynamic models can be directly estimated. The problem of estimating discrete-time models which do not have a continuous-time counterpart is also discussed and some possible causes of this effect are investigated.

This thesis is a contribution to the further application of multisine signals to the measurement and identification of linear and nonlinear systems. It also illustrates the potential of frequency-domain techniques for modelling gas turbine dynamics, where a physical interpretation of the model parameters is to be made.

Table of Contents

Operators, Symbols and Abbreviations

I. Operators

$\{\bullet\}^*$: Complex conjugate of $\{\bullet\}$.

$u * y$: Convolution of u with y.

$E\{\bullet\}$: Expected value.

$\{\bullet\}^{-1}$: Inverse.

$\max\{\bullet\}$: Maximum value.

$\Re\{\bullet\}$: Real part.

$\{\bullet\}^T$: Transpose.

II. Symbols

$\mathbf{a}$: Vector of harmonic amplitudes.

$\mathbf{C}$: Covariance matrix of estimated parameters, equation (A-6).

$D(s)$: Denominator of s-domain transfer function.

$e(t)$: Zero mean white noise sequence.

E : Complex error, equation (4-2).

f : Cyclic frequency (Hz).

$f_{\max}$: Maximum frequency of a band-limited signal.

f_s : Sampling frequency.

f_0 : Fundamental frequency.

F : Number of cosines.

$G(z), H(z)$: Discrete linear transfer functions.

$G_{UU}(\omega)$: Auto-spectrum of $u(t)$.

$G_{YU}(\omega)$: Cross-spectrum of $y(t)$ and $u(t)$.
$H(s)$: Continuous linear transfer function.
$\hat{H}(s)$: Estimate of $H(s)$.
$H(j\omega)$: Linear frequency response function.
$\hat{H}_{EV}(j\omega)$: Error-in-variables estimate of $H(j\omega)$, equation (2-31).
$H_n(s_1, \ldots, s_n)$: nth-order continuous transfer function, equation (6-4).
$H_n(j\omega_1, \ldots, j\omega_n)$: nth-order frequency response function.
$\hat{H}_n(j\omega_1, \ldots, j\omega_n)$: Estimate of $H_n(j\omega_1, \ldots, j\omega_n)$.

$\mathbf{i}$: Vector of input harmonics, equation (3-8).
$K(\mathbf{p})$: Cost function of frequency-domain estimator, equation (2-37).
$K_{\min}$: Expected value of $K(\mathbf{p})$, equation (2-39).
M : Number of averages.
$M(j\omega)$: Input noise spectrum, EV model.
n_p : Number of free parameters.
N : Number of time-domain samples (and MLBS bits in Chapter 3).
$N(s)$: Numerator of s-domain transfer function.
$N(j\omega)$: Output noise spectrum, EV model.
$\mathbf{p}$: Vector of model parameters.
$\hat{\mathbf{p}}$: Vector of estimated parameters.
$\mathbf{p}_0$: Vector of true parameters.
$R_{ee}(j\omega)$: Auto-correlation of frequency-domain residuals, equation (5-15).
$S_n(j\omega)$: Scaling factor for nth-order nonlinearity, equation (5-8).

T_d : Pure time delay.
T_s : Sampling interval.
T_{set} : System settling time.
T_{test} : Total test time.
T_0 : Fundamental period.
$u(t)$: Time-domain input signal.
$U(j\omega)$: True input signal spectrum.
$U_m(j\omega)$: Measured input signal spectrum.
$V_N(\mathbf{p})$: Cost function of time-domain estimator, equation (2-20).
$y(t)$: Time-domain output signal.
$Y(j\omega)$: True output signal spectrum.
$Y_m(j\omega)$: Measured output signal spectrum.
$Y_n(j\omega)$: Spectral output of an n-th order nonlinearity, equation (6-12).

$\gamma^2(\omega)$: Ordinary coherence, equation (2-29).

$\gamma^2_{nl}(\omega)$: Nonlinear coherence, equation (4-15).

$\Delta\mathbf{i}$: Vector of harmonic differences, equation (6-30)

Δt : MLBS bit period.

$\varepsilon(t,\mathbf{p})$: Prediction errors, equation (2-18).

λ : Variance of noise sequence $e(t)$.

$\sigma_{\Delta i}$: Standard deviation of vector of harmonic differences.

$\sigma^2_H(\omega)$: Complex variance of estimated FRF.

$\sigma^2_M(\omega)$: Variance of real or imaginary part of $M(j\omega)$, half the value of equation (3-25)

$\sigma^2_N(\omega)$: Variance of real or imaginary part of $N(j\omega)$, half the value of equation (3-25).

$\sigma_{MN}(j\omega)$: Covariance of $M(j\omega)$ and $N(j\omega)$, half the value of equation (3-26).

τ : Time constant.

Φ : Vector of harmonic phases.

ω : Angular frequency (rad/s).

ω_0 : Fundamental angular frequency.

ω_k : Angular frequency of the k-th component.

III. Abbreviations

AA : Anti-aliasing.

BL : Band-limited.

CF : Crest factor, equation (3-2).

DAC : Digital-to-analogue converter.

dc : Direct current.

EV : Error-in-variables.

FFT : Fast Fourier transform.

FRF : Frequency response function, equation (2-11).

HOFRF : Higher-order frequency response function.

IR : Inverse repeat.

LTI : Linear time-invariant.

MLBS : Maximum length binary sequence.

MLE : Maximum likelihood estimator.

NID : No interharmonic distortion.

NSND : No-sum no-difference.

RMS : Root mean square.

SI : Sparsity index.
SISO : Single-input single-output.
SNR : Signal-to-noise ratio.
TF : Time factor, equation (3-5).
ZOH : Zero-order hold.

Acknowledgments

I would like to extend my thanks to David Rees, my director of studies, for encouraging me to embark on this research project and for supporting me throughout. Special thanks go to Johan Schoukens and Patrick Guillaume, of the Vrije Universiteit Brussel, and István Kollár, of the Technical University of Budapest, for their stimulating discussions and invaluable advice, over several years.

Thanks also to my fellow researchers, Michael Weiss and Antoni Borrell, for many interesting discussions and much practical assistance. I am grateful to the staff of Rolls Royce Derby and DERA Pyestock for the opportunity to work with them on the testing and modelling of an aircraft gas turbine.

I also gratefully acknowledge the following sources:

Figure 2-1 was originally published in *Measurement* and is reproduced by permission of the IMEKO Secretariat.

Figures 7-2, 7-6, 7-8 and 7-9 and Table 7-1 were originally published in the *Advisory Group for Aerospace Research and Development (AGARD) Conference Proceedings No. 324* and are reproduced by permission of AGARD and the respective authors.

Figure 7-3 was originally published in the *AGARD Lecture Series No. 183* and is reproduced by permission of AGARD and the author.

Figure 7-4 was originally published in the *Proceedings of the Institution of Mechanical Engineers* and is reproduced by permission of the Council of the Institution of Mechanical Engineers.

Figure 8-1 was originally published by D.C. Hill in his Ph.D. dissertation and is reproduced by permission of the author.

Chapter I

Introduction

> "The basic reason for the abundant growth of interest in identification is linked to the greatness and the limitations of the human mind. Its greatness, *vis à vis* its ability to grasp situations where dynamic, cause-effect relations are playing an essential role. Its limitations, *vis à vis* the fact that such understanding can only be in terms of a "model", a projection of that part of physical reality which is under consideration." (Eykhoff, 1981)

This thesis deals with the application of modern system identification techniques to engineering problems. Such techniques allow system models to be directly estimated from measured data. The design of the data gathering experiment is a key aspect of any system identification procedure, since the quality of any model directly depends on the quality of the data from which it is estimated. This is the main focus of this thesis, which examines the design of test signals to measure and identify both linear and nonlinear systems.

The thesis can be divided into two main parts; the first part deals with certain theoretical aspects of test signal design and the second part presents the results of testing and modelling a practical system. The design of test signals to identify linear systems is addressed first, in terms of their susceptibility to both noise and to nonlinear effects. This is dealt with in Chapters 3 to 5. Chapter 6 deals with the design of signals to measure points on the frequency-domain Volterra kernels of a nonlinear system.

This work was motivated by the desire to take full advantage of the potential of periodic multifrequency signals. Recent advances in the speed and memory capacity of computer systems mean that such signals can now be easily generated. Multisines are the signals chosen for detailed study, since they allow an arbitrary input spectrum to be defined. This feature is fully exploited in the design approaches presented in this work.

The practical work presented in the second part of the thesis deals with the testing and modelling of an aircraft gas turbine. It will be seen that such a problem is particularly suited to the application of multisine test signals and frequency-domain identification techniques. Chapter 7 serves to introduce this part of the thesis, by presenting a detailed study of the work conducted by other authors in this area, while Chapters 8 and 9 detail

the results of the present study. The overall aim of this work was to estimate models directly from engine data, in order to check, or *verify*, the linearised thermodynamic engine models derived from physics.

The main body of the thesis begins in Chapter 2 with an introduction to linear system theory and system identification techniques. It is here that the basic modelling and identification tools used throughout the thesis are described. The popular prediction error method for discrete time-domain estimation is presented, along with the less well known frequency-domain estimator developed by Schoukens *et al.* (1988). It is shown that each of these estimators exhibits the desirable maximum likelihood properties in the presence of Gaussian noise. Chapter 2 concludes with a discussion of the relative merits of discrete-time and continuous-time models and their application. In applications where an exact representation of the continuous system is required, such as the gas turbine modelling dealt with later in this thesis, then continuous s-domain models should be identified. Since s-domain models are more easily estimated in the frequency domain the main emphasis of this thesis is on frequency-domain analysis and methods.

Chapter 3 deals with the design of the identification experiment. Emphasis is placed on the use of periodic signals and the properties of single sines, multisines and maximum length binary sequences are discussed. A multisine signal is simply an arbitrary sum of harmonically related cosines. The crest factor of a multisine must be minimised if it is to be used for practical testing and a review of the available techniques is presented. This lays the basis for a detailed study of the relative test times using single sine and multisine signals, taking into account both the measurement accuracy and the settling times involved. It is shown that multisine tests can be made shorter than those with single sines and that the major benefit is derived from the reduction in settling time. A frequency-domain approach to the design of maximum length binary sequences is also proposed.

The application of multifrequency signals to nonlinear systems is dealt with in Chapter 4. A review of *block-oriented* nonlinear models is presented, which are composed of various interconnections of static nonlinear and dynamic linear elements. A novel methodology is then presented for studying the frequency-domain contributions generated by both static and dynamic nonlinearities. The contributions are divided into two types, the first of which cannot be influenced by the choice of harmonics included in the signal and the

second of which depends entirely on the choice of harmonics. It is shown that the new analysis method provides a far better insight into the harmonics generated by nonlinearities than previous approaches.

A class of multisine signals containing only odd number harmonics is defined, which include spaces in their spectra for the detection of nonlinear effects. Signals are then designed which do not generate any of the second type of nonlinear contributions at the test frequencies, for a given order of nonlinearity. These signals are termed *no interharmonic distortion* multisines and it is shown that they can be used to measure the best linear approximation of a nonlinear system. This approximation is dependent on the total power of the input signal and a given result will be specific to that signal and the input amplitude used.

Chapter 5 deals with the effects of nonlinearities on multisines, with the aim of designing signals for linear system identification which are robust to these effects. The aim is thus to isolate the linear component of the system response and not, as in Chapter 4, to obtain the best linear approximation of the overall nonlinear system. It is assumed that the systems under study have underlying linear dynamics which dominate the system response at low input signal amplitudes. The influence of the nonlinearity on the test signals is then minimised, for a given input power constraint. The linear dynamics identified in this way should be independent of the input signal and its amplitude. It is shown that the influence of even-order nonlinearities at the test frequencies can be easily eliminated, simply by excluding all the even harmonics from the input signal. The influence of odd-order nonlinearities cannot be eliminated in this way but can be influenced by the harmonics which are included in the signal.

It is found that increasing the number of excluded harmonics, termed the signal *sparsity*, is not the best way to reduce the nonlinear influence at the test frequencies. The key factor in this regard is the signal crest factor. It is shown that a nonlinear error is inherent with any odd-order nonlinearity and that a compensation technique must be employed if it is to be completely eliminated. A method based on testing the system at two signal amplitudes is described, which has the advantage that it can be applied without the need to specify a particular nonlinear model. An overall identification scheme is then proposed at the end of Chapter 5, for identifying linear models in the presence of nonlinear distortions. The systematic application of this approach will ensure that high quality linear models are identified, from which the influence of nonlinearities has been excluded.

The Volterra representation of nonlinear systems is a natural and attractive extension of linear system theory to the nonlinear case, the properties of which have been widely studied. A range of multisine signals is defined in Chapter 6 for directly measuring points on frequency-domain Volterra kernels, which are also known as higher-order frequency response functions. A detailed study of previously designed signals of this type is presented and their drawbacks are discussed. The new design method overcomes many of these problems, by allowing the maximum possible number of kernel points to be measured, maintaining a near-even harmonic spacing and allowing both second- and third-order kernels to be measured simultaneously.

All kernel measurement techniques share a common problem, in that points along certain kernel diagonals cannot be directly measured. A solution to this problem is presented in Chapter 6, using signals made up of paired harmonics, which allow measurements close to the unmeasured diagonals. The chapter ends with an example of the practical application of these signals, which are used to measure the second-order Volterra kernel of a direct current servo motor system.

The second part of the thesis commences with Chapter 7, which is an overview of the work conducted on the dynamic modelling of gas turbines from the early 1950's to the present day. It is clear from this overview that the systematic application of modern system identification techniques to this problem is still in its infancy. At the same time, there is a real need for improved gas turbine models which have been verified against the actual engine performance.

The application of multifrequency test signals has been motivated by the need to reduce engine test times, in order to make systematic engine testing a cost-effective option. The multifrequency signals applied in previous work were maximum length binary sequences and multisines made up of only a small number of harmonics. No attempt had yet been made to systematically assess the influence of the gas turbine nonlinearities on the test signals and consequently on the estimated linear models.

Chapter 7 concludes by defining the key aspects of this current investigation, which include: applying wide-band, low crest factor multisine signals to the testing of gas turbines; directly estimating s-domain models using frequency-domain techniques; systematically assessing the influence of system nonlinearities; and estimating the engine pure time delay. The aim was to estimate engine models, along with a measure of their accuracy, in order to verify the linearised thermodynamic models derived from physics.

The experimental work was conducted on a Rolls Royce Spey Mk 202 engine, at the Glen sea-level test facility of the Defence Evaluation and Research Agency at Pyestock. This engine is a typical military twin-shaft turbofan and the work focussed on modelling the fuel feed to shaft speed dynamics. Chapter 8 deals with the design of appropriate test signals for the engine, in order to perform linear modelling and cross-validation and also nonlinear detection and modelling. The signals described were applied to the engine and a detailed analysis of the results is presented. The advantages of using periodic multisine signals are clearly illustrated, in both the data analysis and the subsequent modelling.

Engine frequency responses are calculated for each of the tests, along with uncertainty bounds, which show that the uncertainty on the estimates is very small. It is concluded that broad-band multisine signals can be applied with confidence to engine testing, in place of the more commonly used single sine tests, thus providing a considerable reduction in test times and consequently reducing costs.

Chapter 9 deals with the frequency-domain identification of s-domain engine models using multisine data. High quality models of each shaft are estimated and cross-validated using data gathered using a different test signal. The pure time delay is included as a parameter for estimation and the results are close to those obtained in previous studies for a typical gas turbine. A weak even-order nonlinearity is detected in the engine, which does not affect the odd-harmonic multisines used. A comparison of the model dynamics shows that the shaft dynamics appear to be faster than predicted by the thermodynamic models and that the second-order dynamics of the low pressure shaft are also more significant than predicted.

It will thus be seen that frequency-domain techniques can be used to estimate linear engine models which have low uncertainty and are free from the influence of significant nonlinear effects. The use of odd harmonic multisines allows the linear and nonlinear effects to be clearly separated and the shaft dynamics to be accurately identified.

The time-domain estimation of discrete-time models is also studied for comparison, using a prediction error approach, and models with good input-output properties are obtained. The problem is that the preferred models for each shaft contain a single real negative pole. Such a pole has no continuous counterpart and cannot be compared to the poles of the s-domain thermodynamic models. The possible sources of this error are discussed and studied through simulation.

The problem is eliminated by resampling the engine data at a faster rate. The use of periodic inputs signals allows the data to be resampled using the fast Fourier transform and its inverse. It is thus possible to estimate discrete-time engine models which are equivalent to continuous-time models but the results strongly depend on the sampling frequency selected. It is concluded that the estimation of discrete-time models is not the best approach for this specific problem, that of verifying the thermodynamic engine models. The work presented in Chapters 8 and 9 illustrates the potential of frequency-domain techniques for modelling industrial systems where a physical interpretation of the model parameters is to be made.

The chapters are written to be self-contained, as far as possible, and each begins with a short abstract and ends with a set of conclusions. As such, both Chapter 3 and Chapter 6 can be read on their own, without reference to the other work. It is suggested that Chapters 4 and 5 should be read together, as should Chapters 7 to 9, which form a coherent sub-section of the thesis.

A number of the chapters are based on one or more published papers. Chapters 4 and 5 are based on Evans *et al.* (1992, 1994a) and on the first part of Evans *et al.* (1995). Chapter 6 is a revised and updated version of Evans *et al.* (1996), which was an invited tutorial paper. The work on gas turbines described in Chapters 8 and 9 has been presented at a number of conferences (Evans *et al.*, 1994b and 1994c) and a journal paper discussing the results has been accepted for publication (Evans *et al.*, 1998). The gas turbine results were also included as an application example in Evans *et al.* (1995) and this paper was awarded the 1996 F.C. Williams Premium by the Institution of Electrical Engineers, which is the highest award for a published paper presented by its Computing and Control Division.

Chapter II

Linear System Identification

Abstract —*An introduction is provided to the fundamental concepts relating to the modelling and estimation of linear time invariant systems in the time and frequency domains. The role of system modelling is discussed and a review of fundamental linear modelling concepts is presented. Estimators are described for parametric time-domain modelling and for both parametric and nonparametric frequency-domain modelling. A comparison is made between discrete-time and continuous-time models and their appropriate applications are discussed. No original results are presented in this chapter and the techniques described are well established estimation tools, which will be employed throughout this thesis.*

2.1 Introduction

Modern system identification techniques offer a powerful method of projecting some aspects of physical reality into a mathematical description or *model.* While a relatively new field, the area has received much attention over the last thirty years and a mature body of knowledge and tested techniques now exist (Eykhoff, 1984).

The predominant ideas of the 19th century saw science and engineering as a process of *discovering* the underlying laws which governed the behaviour of physical systems. Since that time, the emphasis has gradually moved towards that of *developing,* or more recently *identifying*, appropriate models. This change has been motivated by the realisation that models can only describe certain aspects of the real world and that any mathematical model can only be a partial representation of complex reality. The success of the modelling effort must then be judged by how useful the final model proves to be in its intended application.

A distinction must consequently be made between the *system* and the *model.* The system refers to the actual real-life system, whereas the model refers to a necessarily limited mathematical description of that system. Eykhoff (1981) offered a concise definition of a model as: "a representation of essential aspects of a system, which offers the knowledge of that system in a useful form."

According to Fasol and Jörgl (1980), the use of system identification can be grouped into six main areas, the first two of which are of particular relevance to the work described later in this thesis.

- Obtaining a better knowledge of the process;
- Verification of theoretical models;
- Synthesis of control systems;
- Prediction of signals or system outputs;
- Optimisation of process behaviour;
- Computation of variables which cannot be directly measured.

A model can be expressed in *parametric* form, using a finite number of parameters, usually of differential or difference equations. Alternatively, a model can be expressed in *nonparametric* form, as a set of values related to an independent variable, usually time or frequency. In theory, an infinite set of values is required to completely characterise a system using the nonparametric approach but in practice the values are usually restricted to a finite time or frequency range.

As mentioned above, the classical approach to deriving a parametric model is to develop theoretical *laws* which corresponded to the physical characteristics of the system in question. Much of modern engineering was elaborated in this way, using the basic laws of Newtonian mechanics. Such models have proved their usefulness in numerous applications, such as discovering planets, building bridges, designing engines and the development of propelled flight. They continue to find widespread application to this day.

Modern technology now provides an alternative method of deriving a system model, that of *estimating* it from values of the system input and corresponding output. The parameters of such a model do not necessarily correspond to any physical characteristics of the system. This approach has only become generally applicable since the advent of digital computers, although the underlying mathematical techniques were developed in the early nineteenth century by Gauss (Åström, 1980).

The two paths are illustrated in Figure 2-1, which is the scheme presented by Eykhoff (1984) in an excellent survey paper on this subject. It should be stressed that any practical parametric model building will involve elements of both paths and that models estimated from input and output data can often be used to verify the theoretical models derived from physics.

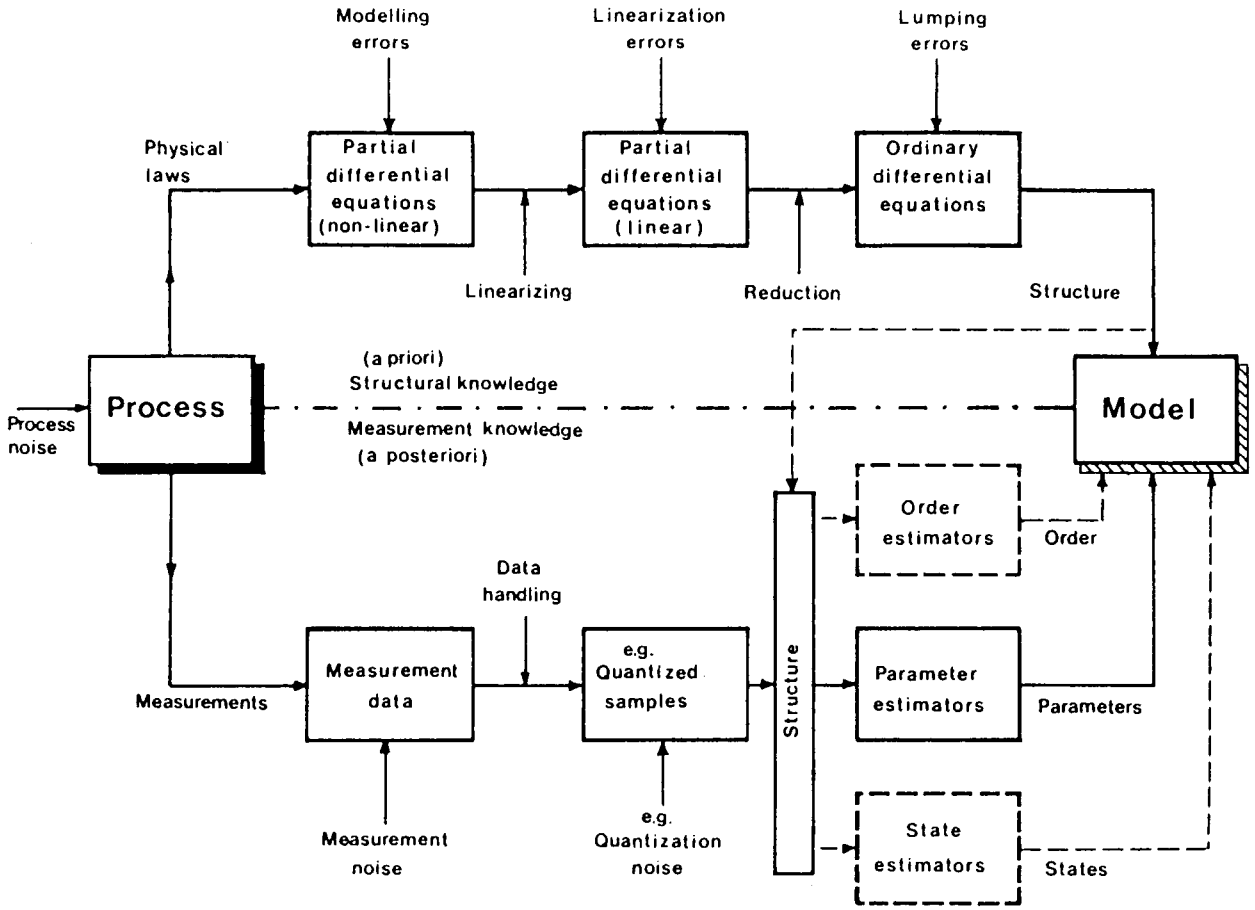

Figure 2-1. A parametric identification scheme.
(Reproduced from Eykhoff (1984), with permission.)

2.2 Linear Models

The basic relationship between the input and output of a causal *linear time-invariant* (LTI), *single-input single-output* (SISO) system is given by the convolution or superposition integral

$$y(t) = \int_{\tau=0}^{t} h(\tau)\, u(t-\tau)\, d\tau \qquad t \geq 0 \tag{2-1}$$

where $u(t)$ is the independent variable, $y(t)$ the dependent variable and $h(\tau)$ is termed the system unit impulse response. It is possible to calculate the output $y(t)$, if $h(\tau)$ and $u(t)$ are known across the appropriate limits of integration. Hence, the impulse response is a complete nonparametric time-domain characterisation of an LTI system.

Since most basic physical relationships are expressed in terms of differential equations it is natural to work in the continuous-time domain. Taking the Laplace transform of equation (2-1) gives an input-output relationship in terms of the complex variable s

$$Y(s) = H(s)\, U(s) \tag{2-2}$$

where $Y(s)$ and $U(s)$ are the Laplace transforms of $y(t)$ and $u(t)$ and $H(s)$ is termed the continuous transfer function. This can be expressed as a rational polynomial function in s, with the number of zeros less than or equal to the number of poles ($nb \leq na$).

$$H(s) = \frac{b_0 + b_1 s + \ldots b_{nb} s^{nb}}{a_0 + a_1 s + \ldots a_{na} s^{na}} = \frac{N(s)}{D(s)} \tag{2-3}$$

A further restriction that the transfer function must be strictly proper ($nb < na$) implies that there is no direct term relating the system input and output. In practice, the system output will be observed at discrete sampling instants

$$y(kT_s) = \int_{\tau=0}^{t} h(\tau)\, u(kT_s - \tau)\, d\tau \qquad k = 1, 2, \ldots \tag{2-4}$$

where T_s is the sampling interval. If the input signal is kept constant between sampling

instants, termed *piece-wise constant*, such that

$$u(t) = u(kT_s) \qquad kT_s \leq t < (k+1)T_s \tag{2-5}$$

then the convolution integral of equation (2-1) can be replaced, without approximation, by a discrete summation

$$y(kT_s) = \sum_{l=1}^{t} h_{T_s}(l)\, u(kT_s - l) \qquad \text{where} \quad h_{T_s}(l) = \int_{\tau=(l-1)T_s}^{lT_s} h(\tau)\, d\tau \tag{2-6}$$

which defines the system output at the discrete time points kT_s (Ljung, 1987, p. 14). The term $h_{T_s}(l)$ is the discrete impulse response of the system. There is no approximation involved in the discretisation if it is assumed that a *zero-order-hold* (ZOH) device is used to apply the signal to the continuous system. This is termed the *ZOH assumption*, which will be discussed in greater detail in Chapter 3.

Defining a backward shift operator z^{-1}, where

$$z^{-1}\, u(kT_s) = u((k-1)T_s) \tag{2-7}$$

then equation (2-6) can be re-written as

$$y(kT_s) = \left[\sum_{l=1}^{t} h_{T_s}(l)\, z^{-l} \right] u(kT_s) = H_{T_s}(z)\, u(kT_s) \tag{2-8}$$

where $H_{T_s}(z)$ can be expressed in closed form as a rational polynomial function in z, termed the discrete transfer function

$$H_{T_s}(z) = \frac{b_1 z^{-1} + \ldots + b_{nb} z^{-nb}}{1 + a_1 z^{-1} + \ldots + a_{na} z^{-na}} = \frac{B(z)}{A(z)} \tag{2-9}$$

The subscript T_s in equations (2-6) and (2-9) indicates that the parameters of the discrete models are dependent on the sampling interval. The exclusion of the coefficient b_0 in the numerator of $H_{T_s}(z)$ means that models with a direct term, whereby the input instantly

affects the output, will not be considered. This implies that a change in the input to the ZOH at a sampling instant kT_s will not have any instantaneous effect on the system output and will only be reflected in the output at the next sampling instant $(k+1)T_s$, or later.

A continuous transfer function $H(s)$ can be transformed to a discrete transfer function $H_{T_s}(z)$ without approximation, under the ZOH assumption. The *impulse invariant transform* is defined such that the impulse response of the discrete system is equal to that of the underlying continuous system at the sampling instants. The mapping of the continuous poles is defined by the straightforward relation

$$z = e^{sT_s} \qquad (2-10)$$

but the mapping of the continuous zeros is considerably more complex, as discussed by Åström *et al.* (1984). The problem of transforming the continuous zeros can be avoided by expressing the transfer function in terms of partial fractions, as modes and modal gains.

Information about the continuous system will be lost if the discrete sampling frequency is too low, in a similar fashion to the problem of aliasing with sampled signals. If the sampling frequency is less than twice the break-point frequencies of the continuous poles then aliased poles will appear in the z-domain, at frequencies lower than their true values.

Sampling too quickly, on the other hand, can lead to numerical problems, since as T_s approaches zero in equation (2-10) the z-domain poles will converge to a single point at $z = 1$, irrespective of the values of s. Faster sampling can also give rise to non-minimum phase discrete models (with zeros outside the unit circle), of minimum phase continuous systems (with all the zeros in the left-hand s-plane).

A *pure time delay*, sometimes known as *dead time*, can be included in a continuous model by multiplying $H(s)$ by a term e^{-sT_d}, where T_d is the time delay. A delay is incorporated into a discrete model by delaying the input by a certain number of sampling intervals. This means that the time delay is restricted to integer multiples of the sampling interval for discrete systems and shows up as zero values of the leading numerator coefficients.

2.2.1 Frequency Domain

The basic input-output relationship in the frequency domain is given by the Fourier transform of the convolution integral in equation (2-1). Under steady-state conditions, where all transient effects have died away, the relation reduces to a simple multiplication

$$Y(j\omega) = H(j\omega)\, U(j\omega) \qquad (2-11)$$

where $Y(j\omega)$ and $U(j\omega)$ are the Fourier transforms of $y(t)$ and $u(t)$ and $H(j\omega)$ is a complex valued function of frequency, termed the *frequency response function* (FRF) of the system.

The steady state frequency response of a continuous model $H(s)$ can be evaluated by substituting $j\omega$ for s in equation (2-3) and evaluating the polynomials along the imaginary axis. In a similar fashion, the frequency response of the discrete model $H_{T_s}(z)$ can be evaluated by substituting $e^{j\omega T_s}$ for z, which involves evaluating the polynomials around the unit circle.

2.3 Model Estimation

The process of model estimation involves three key elements: data gathering, selection of a model set from which to estimate, and selection of a rule by which the most appropriate model can be chosen. Once a model has been chosen a further decision must be made, with regard to the quality of the model for its intended application, termed *validation*. A deficiency in the model may be due to a variety of reasons, which are summarised by Ljung (1987) as:

- The numerical procedure failed to find the best model;
- The selection criterion was not well chosen;
- The model set was not appropriate;
- The data set was not informative enough.

The last two reasons are of particular relevance to the work described in this thesis, since it is concerned with the influence of nonlinear errors on linear system identification and the design of test signals to gather information from the system.

2.3.1 Time Domain

Estimation of the system impulse response, defined in equation (2-1), can be carried out either by subjecting the system to an impulsive input and measuring the output, or by using binary sequences and correlation methods (Godfrey, 1980). The correlation approach was particularly popular during the 1960's and 1970's but its application has declined in recent years, since the increase in computer processing power now allows more sophisticated techniques to be employed.

A general structure for the class of discrete LTI transfer function models is shown in Figure 2-2, which can be expressed as

$$A(z)\,y(t) = \frac{B(z)}{F(z)}u(t) + \frac{C(z)}{D(z)}e(t) \qquad (2-12)$$

where

$$A(z) = 1 + a_1 z^{-1} + \ldots + a_{na} z^{-na} \qquad (2-13)$$

$$B(z) = b_1 z^{-1} + \ldots + b_{nb} z^{-nb} \qquad (2-14)$$

and $F(z)$ has the same form as $A(z)$, while $C(z)$ and $D(z)$ have the same form as $B(z)$. The system input is $u(t)$, its output is $y(t)$ and the noise sequence $e(t)$ is assumed to be normally distributed and independent of the input signal, with zero mean and variance λ_0. Simpler model structures can be achieved by setting the orders of certain polynomials to zero. Two basic model classes exist, the first being *equation error* models, where the noise term enters as a direct error in the difference equations. Models in this class include:

- *Auto Regressive with eXogenous input* (ARX): nc=nd=nf=0, so called because the predictor can be defined as a linear regression.
- *Auto Regressive Moving Average with eXogenous input* (ARMAX): nd=nf=0, where $C(z)$ acts as a moving average filter of the noise.

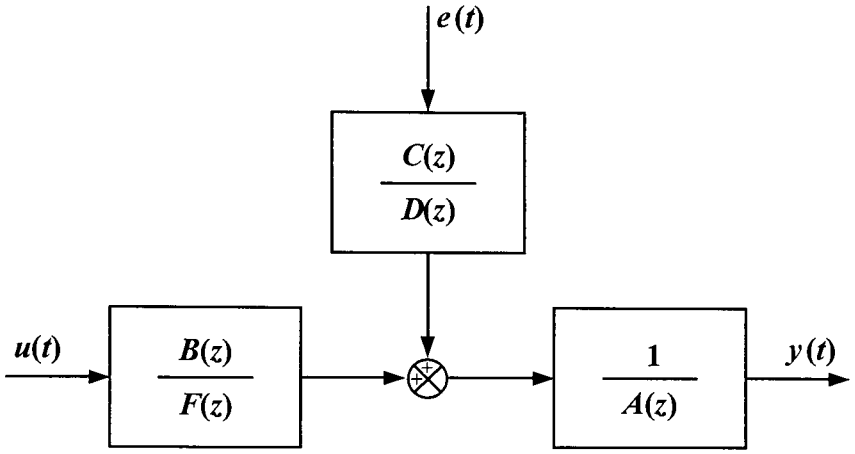

Figure 2-2. A general discrete transfer function model.

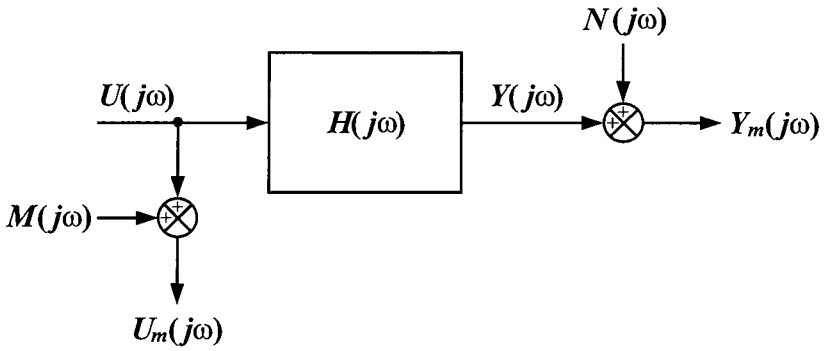

Figure 2-3. An error-in-variables model.

The second class of models are called *output error* structures where, as the name suggests, the noise term is added to the system output. Such models include:

- *Output Error* (OE): na=nc=nd=0.
- *Box-Jenkins* (BJ): na=0; which allows greater flexibility in the choice of the noise model at the cost of increased complexity.

Further details of the different structures which can be employed are given in Wellstead and Zarrop (1991, pp. 57-60) and Ljung (1987, chapter 4). The general model can be compactly expressed as

$$y(t) = G(z,\mathbf{p})\,u(t) + H(z,\mathbf{p})\,e(t) \qquad (2-15)$$

where $G(z,\mathbf{p})$ and $H(z,\mathbf{p})$ are filters of finite order and functions of a parameter vector $\mathbf{p}$.

$$\mathbf{p} = [\, a_1 \ldots a_{na} \quad b_0 \ldots b_{nb} \quad c_1 \ldots c_{nc} \quad d_1 \ldots d_{nd} \quad f_1 \ldots f_{nf} \,]^T \qquad (2-16)$$

The one-step-ahead prediction for equation (2-15) is denoted as $\hat{y}(t)$

$$\hat{y}(t) = H^{-1}(z,\mathbf{p})\,G(z,\mathbf{p})\,u(t) + [\,1 - H^{-1}(z,\mathbf{p})\,]\,y(t) \qquad (2-17)$$

The errors between the actual process output and the predicted model output can be calculated by subtracting equation (2-17) from equation (2-15)

$$\varepsilon(t,\mathbf{p}) = y(t) - \hat{y}(t) = H^{-1}(z,\mathbf{p})\,[\,y(t) - G(z,\mathbf{p})\,u(t)\,] \qquad (2-18)$$

These are variously termed the *residuals, prediction errors* or *innovations* and the family of estimators which use this approach are called *prediction error methods* (PEM). If the filter $H(z)$ is set to unity then the predicted error calculated according to equation (2-18) becomes

$$\varepsilon(t,\mathbf{p}) = y(t) - G(z,\mathbf{p})\,u(t) \qquad (2-19)$$

which is the difference between the measured output $y(t)$ and the noise-free model output, corresponding to the OE structure described above. The most common approach is to

determine the estimates of $G(z)$ and, if appropriate, $H(z)$ by minimising a quadratic cost function.

$$V_N(\mathbf{p}) = \sum_{t=1}^{N} \varepsilon^2(t, \mathbf{p}) \qquad (2-20)$$

$$\hat{\mathbf{p}}_N = \arg\min V_N(\mathbf{p}) \qquad (2-21)$$

Where arg min means the argument which minimises the function and $\hat{\mathbf{p}}_N$ is the estimated parameter vector. An analytical solution can be found to this problem for the ARX model but this is not the case for the other model structures and a nonlinear least squares technique must be used, such as the Gauss-Newton approach (Ljung, 1987, pp. 282-294).

A pure time delay can be modelled by delaying the input by a certain number of sampling intervals. An empirical approach to selecting this delay is suggested by Ljung (1995, p. 3-54), which involves estimating a second-order ARX model and finding the delay which minimises the cost function. This delay is then used for further modelling, using alternative structures and model orders, and once a satisfactory model has been found the delay is once again varied until the best model fit is found.

The properties of PEM estimators have been extensively studied and it has been shown that they are *maximum likelihood estimators* (MLE) in the presence of Gaussian innovations (Ljung, 1987, chapters 8 and 9). Provided that the true model is part of the model set and that the data are informative enough (discussed in Chapter 3) then the estimates will converge in mean square to the true parameter values $\mathbf{p}_0$ (as $N \rightarrow \infty$), so that the estimator is *asymptotically unbiased.* If the true system is not part of the model set then the estimates will converge to the best approximation available in the set.

If the measurements are *independent and identically distributed* (iid) then an MLE is strongly consistent and the covariance matrix of the estimates reaches the Cramér-Rao bound asymptotically, so that the estimator is asymptotically efficient (Norton, 1986, pp. 133-142). The identical distribution of the measurements implies that the disturbing noise has stationary statistical properties over the measurement interval. A summary of the terms and concepts relating to estimation theory is provided in the Appendix at the end of this chapter.

The covariance matrix can be approximated by the Cramér-Rao bound, under the assumption that the estimator has reached its asymptotic properties (N is sufficiently large). This can be estimated from the data as:

$$\mathbf{C}_{CR} = \hat{\lambda}_N \left[\frac{1}{N} \sum_{t=1}^{N} \psi(t,\mathbf{p})\, \psi^T(t,\mathbf{p}) \right]^{-1} \qquad (2-22)$$

which is the inverse of the covariance matrix of the predictor gradients $\psi(t,\mathbf{p})$, normalised by the innovations variance $\hat{\lambda}_N$ (Ljung, 1987, p. 243). The variances of the parameter estimates are found on the diagonal of this matrix. The predictor gradients are defined as

$$\psi(t,\mathbf{p}) = \frac{d\,\hat{y}(t \mid \mathbf{p})}{d\,\mathbf{p}} = -\frac{d\,\varepsilon(t,\mathbf{p})}{d\,\mathbf{p}}\Big|_{\mathbf{p}=\mathbf{p}_0} \qquad (2-23)$$

which is a column vector of the partial derivatives of $\hat{y}(t \mid \mathbf{p})$ with respect to each of the parameters. This should be evaluated at the true parameter values $\mathbf{p}_0$ but since these are unknown the estimated parameters must be used in their place. Since $\psi(t,\mathbf{p})$ is the gradient of $\hat{y}(t \mid \mathbf{p})$ then equation (2-22) clearly implies that the asymptotic accuracy of a given parameter is related to the sensitivity of the prediction with respect to that parameter. The innovations variance can be estimated as

$$\hat{\lambda}_N = \frac{1}{N} \sum_{t=1}^{N} \varepsilon^2(t,\hat{\mathbf{p}}_N) \qquad (2-24)$$

Strictly speaking, the expression in equation (2-22) is only valid if the true model lies within the set of all possible models which can be estimated. This is rarely the case, as was discussed at the very beginning of this chapter. However, the expression can still be considered as a useful approximation of the parameter variances, as long as the modelling errors are not too severe.

The maximum likelihood property also implies that the estimates are asymptotically normally distributed, which allows parameter confidence bounds to be calculated in a straightforward manner. The parameter variances can also be translated into confidence ellipsoids of the poles and zeros, using the linear approximation technique of Gauss. This is particularly useful, since the uncertainties of the model poles and zeros are far easier to

interpret than the uncertainty of individual model parameters. The transformation must be treated with some caution however, since the behaviour of estimated poles and zeros across several experiments with noisy data cannot always be modelled in a linear fashion.

The estimators and algorithms discussed have been implemented as part of the Matlab System Identification Toolbox, written by Ljung (1995).

2.3.2 Frequency Domain

A general *error-in-variables* (EV) model for frequency-domain identification of LTI systems is shown in Figure 2-3, where the measured Fourier coefficients of the input and output $U_m(j\omega)$ and $Y_m(j\omega)$, are corrupted by noises $M(j\omega)$ and $N(j\omega)$. It will be assumed that the noise sequences have the following properties:

- They have a complex normal distribution, which means that the real and imaginary parts of their Fourier coefficients are independent and normally distributed, with an equal variance (Ljung, 1987, p. 459).
- The sequences have zero means, $E\{ M(j\omega) \} = E\{ N(j\omega) \} = 0$
- The noise is uncorrelated between frequency points, such that $E\{ M(j\omega_i) M(j\omega_k) \} = 0$ for $i \neq k$, and similarly for $N(j\omega_k)$.
- The noises are independent of the input Fourier coefficients.

Brillinger (1975, pp. 88-116) has shown that, under mild conditions (which include stationarity), the frequency-domain noise observed via the discrete Fourier transform will approach these properties asymptotically, as the number of time samples increases. It has also been shown that the asymptotic properties are well approximated for values of N as small as 512 samples (Schoukens and Renneboog, 1986). A further assumption, that the input and output noises are independent, will also be made for certain derivations in this thesis. This will be explicitly stated where appropriate.

Most commercially available frequency response analysers calculate the nonparametric FRF by exciting the system with a random signal and estimating the frequency response from ratios of the cross- and auto-spectra of the input and output signals. The two approaches most commonly implemented are the classical H_1 and H_2 estimators

$$H_1(j\omega) = \frac{\frac{1}{M}\sum_{m=1}^{M} U_m^*(j\omega)\, Y_m(j\omega)}{\frac{1}{M}\sum_{m=1}^{M} U_m^*(j\omega)\, U_m(j\omega)} = \frac{G_{UY}(j\omega)}{G_{UU}(\omega)} \qquad (2-25)$$

$$H_2(j\omega) = \frac{\frac{1}{M}\sum_{m=1}^{M} Y_m^*(j\omega)\, Y_m(j\omega)}{\frac{1}{M}\sum_{m=1}^{M} Y_m^*(j\omega)\, U_m(j\omega)} = \frac{G_{YY}(\omega)}{G_{YU}(j\omega)} \qquad (2-26)$$

where $U_m(j\omega)$ and $Y_m(j\omega)$ are measured across M independent time blocks of equal length. Time-domain windowing is normally employed if a random excitation signal is used, in order to minimise the leakage errors introduced by the *fast Fourier transform* (FFT), a topic which has been thoroughly reviewed by Gade and Herlufsen (1987).

Assuming that the input and output noises are uncorrelated with each other and with the input and output signals, the following results can be obtained

$$H_1(j\omega) = H(j\omega)\frac{1}{1 + [G_{MM}(\omega) / G_{UU}(\omega)]} \qquad (2-27)$$

$$H_2(j\omega) = H(j\omega)\left(1 + \frac{G_{NN}(\omega)}{G_{YY}(\omega)}\right) \qquad (2-28)$$

where $G_{MM}(\omega)$ and $G_{NN}(\omega)$ are the auto-spectra of the input and output noise. The equations show that the H_1 estimator is biased, in magnitude only, in the presence of input noise and H_2 is similarly biased for noise on the output. The bias depends on the *signal-to-noise ratio* (SNR) at the input and output respectively, and hence on the second-order moments (i.e. power) of the noise (Bendat and Piersol, 1980, p. 89).

The quality of the FRF estimates is usually assessed by calculating the *squared coherence function*, defined as

$$\gamma^2(\omega) = \frac{|\, G_{UY}(j\omega)\,|^2}{G_{UU}(\omega)\, G_{YY}(\omega)} \qquad (2-29)$$

which is the H_1 estimator divided by the H_2 estimator. It can be shown that

$$| G_{UY}(j\omega) |^2 \leq G_{UU}(\omega)\, G_{YY}(\omega) \qquad (2-30)$$

and hence the coherence can vary between zero and one, where a value of one indicates an identical linear relationship between $U_m(j\omega)$ and $Y_m(j\omega)$ for each of the measurements (Bendat and Piersol, 1980, p. 56). Since the cross- and auto-spectra are averaged quantities the coherence can be thought of as a measure of the deterministic nature of the system frequency response. If the estimated frequency response is constant for each measurement the coherence will be one. If random signals are used to excite the system then coherence values of less than one will be caused by:

- Uncorrelated noise on the measurements;
- Nonlinearities in the system under investigation;
- Leakage errors in the FFT.

Each of these effects will produce a random (stochastic) influence on the estimated response, the first due to its inherent nature and the second and third due to the nature of the input signal. Thus, when random signals are used, the coherence can be interpreted as a measure of the fraction of the output spectrum which is linearly related to the input spectrum at a given frequency.

However, care must be taken in the interpretation of the coherence if periodic input signals are used, with averaging carried out over complete signal periods. The nonlinear contributions at the output will now be deterministic, since the response of the nonlinearity to each signal period will be the same, and there will be no drop in the coherence. Thus, if periodic signals are used, the coherence will only allow the detection of stochastic effects and will not allow the linear and nonlinear behaviour of the system to be distinguished.

The use of periodic input signals allows an alternative estimator to be defined, which is simply the ratio of the mean values of the output and input Fourier coefficients, at the discrete test frequencies ω_k.

$$\hat{H}_{EV}(j\omega_k) = \frac{\frac{1}{M}\sum_{m=1}^{M} Y_m(j\omega_k)}{\frac{1}{M}\sum_{m=1}^{M} U_m(j\omega_k)} = \frac{\overline{Y}(j\omega_k)}{\overline{U}(j\omega_k)} \qquad (2-31)$$

This has been termed the EV estimator and it has been shown that it is an MLE if the input and output noises have a complex normal distribution, even if they are mutually correlated (Guillaume, 1992). The bias of the estimator depends on the fourth-order moments of the noise and, with synchronised measurements of the input and output signals, the estimator is both asymptotically unbiased and efficient.

It is also important to consider the uncertainty of the FRF estimates. Schoukens *et al.* (1993a, pp. 153-155) have shown that, using linear approximations, each of the estimators described above has the same uncertainty. Defining the variance of the estimated FRF as

$$\sigma_H^2(\omega) = E\{ (H(j\omega) - E\{H(j\omega)\}) (H(j\omega) - E\{H(j\omega)\})^* \} \qquad (2-32)$$

which can be written as

$$\sigma_H^2(\omega) = \sigma_{|H|}^2(\omega) + |H(j\omega)|^2 \sigma_{\angle H}^2(j\omega) \qquad (2-33)$$

the approximate variance of the FRF estimates is given by:

$$\begin{aligned}\sigma_H^2(\omega_k) &= 2\frac{|H(j\omega_k)|^2}{M}\left(\frac{\sigma_M^2(\omega_k)}{|U(j\omega_k)|^2} + \frac{\sigma_N^2(\omega_k)}{|Y(j\omega_k)|^2}\right) \\ &= \frac{2}{M|U(j\omega_k)|^2}(|H(j\omega)|^2 \sigma_M^2(\omega_k) + \sigma_N^2(\omega_k)) \end{aligned} \qquad (2-34)$$

where the terms $\sigma_M^2(\omega_k)$ and $\sigma_N^2(\omega_k)$ are the variances of the real or imaginary parts of $M(j\omega_k)$ and $N(j\omega_k)$. If the input and output noises are cross-correlated then an additional factor must be introduced, such that

$$\sigma_H^2(\omega_k) = \frac{2}{M|U(j\omega_k)|^2}(|H(j\omega)|^2 \sigma_M^2(\omega_k) + \sigma_N^2(\omega_k) - 2\Re\{\sigma_{MN}(j\omega_k)H^*(j\omega_k)\}) \qquad (2-35)$$

where $\sigma_{MN}(j\omega_k)$ is the noise cross-covariance. Correlated noise may be the result of input noise passing through the system or of the system operating within a feedback loop. It can be seen from equation (2-35) that the variance of the estimated FRF is inversely proportional to the number of measurements and the power of the input harmonics and proportional to the noise variances *referred to the system output.*

Illustrative Example. The amplitude bias of the different estimators in the face of input and output noise can be illustrated through simulation. A system was simulated with a transfer function $H(j\omega)=1$ for all frequencies and a periodic input signal applied, with unity amplitude at each frequency.

Simulated data were created by adding zero mean normally distributed noise to the input and output frequency amplitudes. The noise variances were varied to generate frequency domain SNR's in the range of 0-30 dB. The number of averages M was 1000, in order to illustrate the asymptotic properties of the estimators.

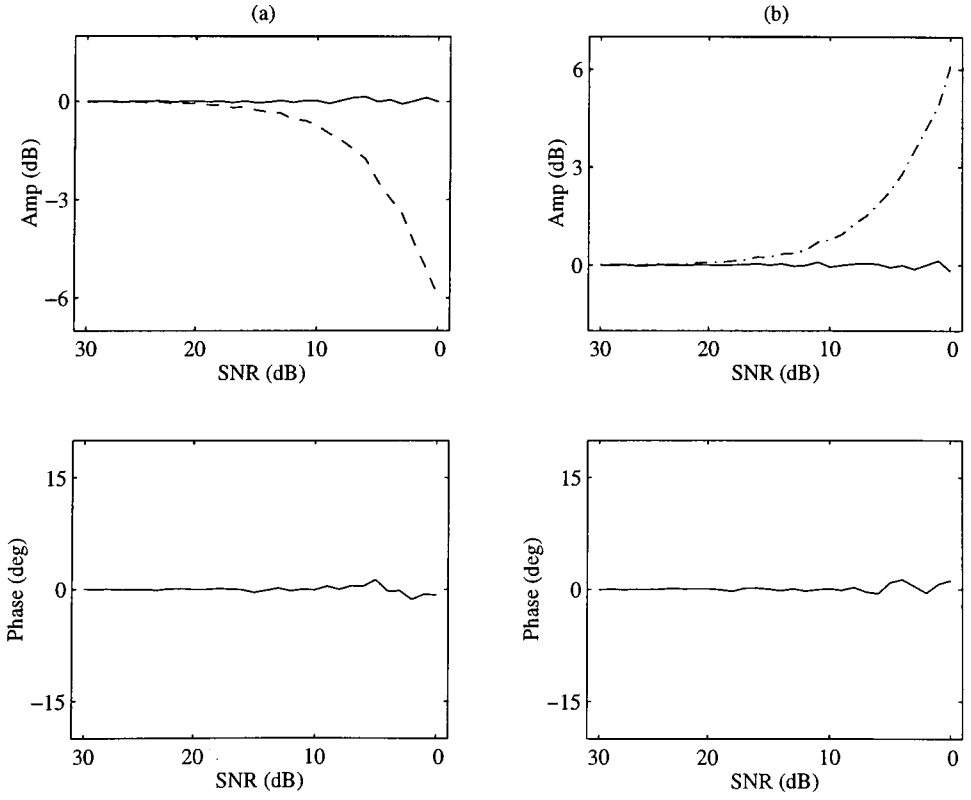

Figure 2-4. Amplitude and phase of FRF estimates with (a) input noise and (b) output noise. $H_{EV}(j\omega)$ (solid) $H_1(j\omega)$ (dashed) and $H_2(j\omega)$ (dash-dot).

The results are plotted in Figure 2-4, which shows the amplitude bias of the H_1 estimator in the presence of input noise and the H_2 estimator in the presence of output noise. For practical purposes, the three estimators can be considered as unbiased for SNR's of greater than 25 dB.

The variation of the uncertainty with SNR can also be illustrated using the simulation example. The 1σ uncertainty bounds for the unbiased H_{EV} estimator are shown in Figure 2-5, which were estimated using equation (2-34).

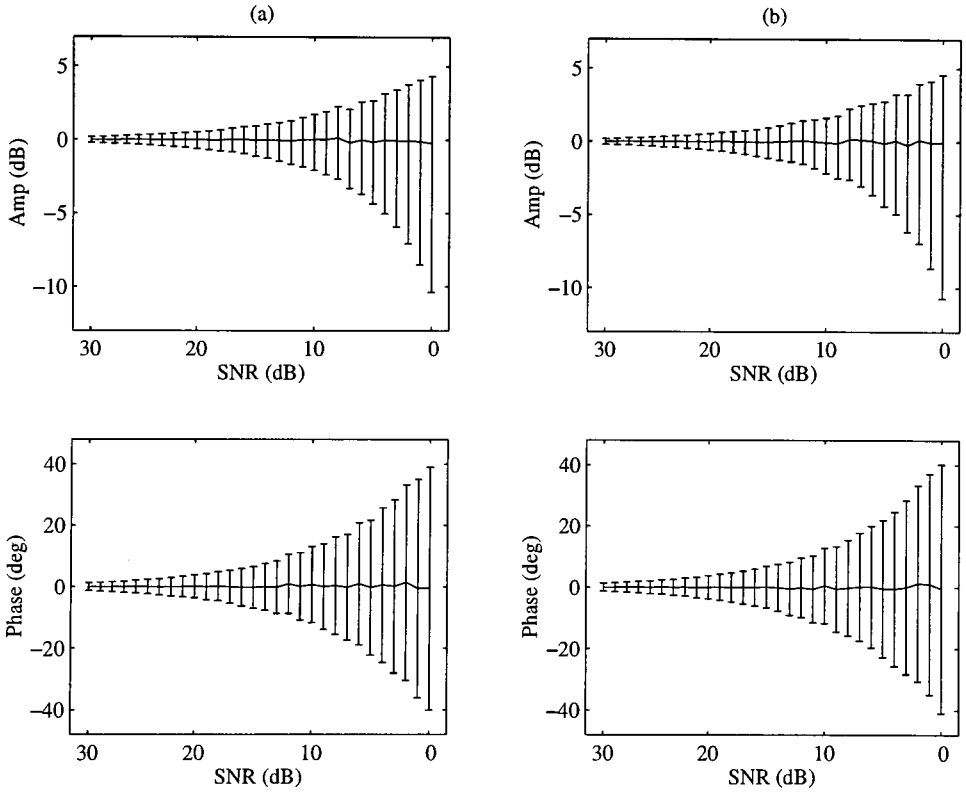

Figure 2-5. Uncertainty of $H_{EV}(j\omega)$ estimates, showing 1σ bounds with (a) input noise and (b) output noise.

It is possible to use the simulation to investigate the validity of the linear approximation used in the derivation of equation (2-35), by comparing the actual standard deviation of the FRF with that estimated using the equation. These bounds are plotted in Figure 2-6, where it can be seen that the approximation is extremely good down to SNR's of 10 dB and remains a useful measure of the uncertainty even as the SNR approaches 0 dB.

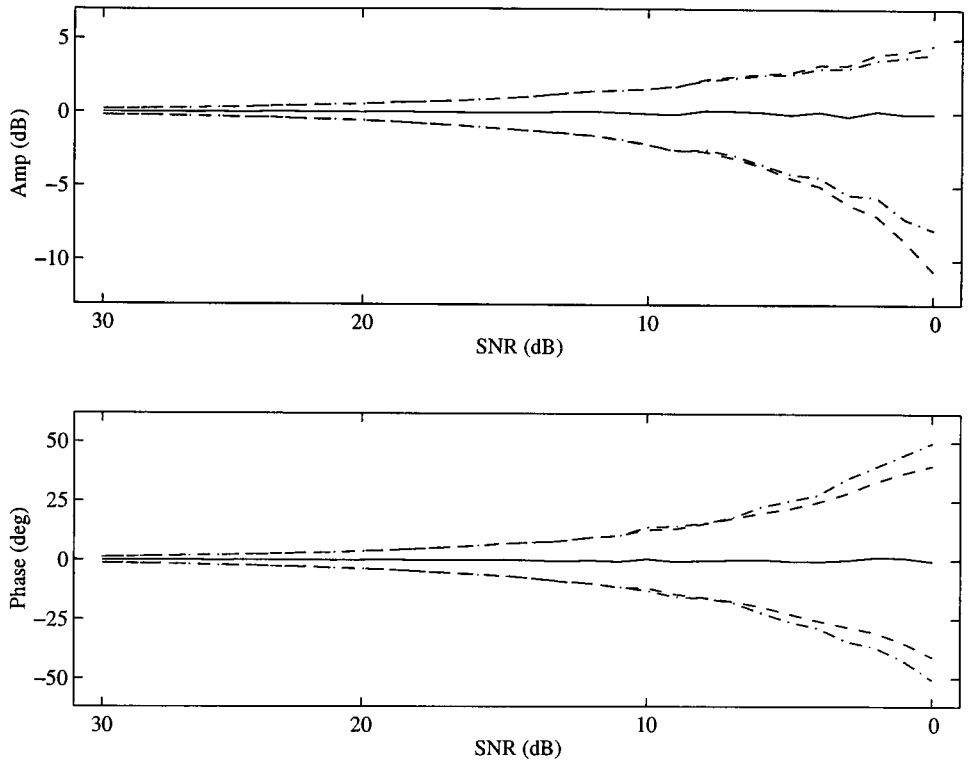

Figure 2-6. Comparison of estimated (dashed) and actual (dash-dot) 1σ bounds.

Parametric frequency-domain identification involves selecting the parameters of an *s*- or *z*-domain model, with pure time delay T_d

$$H(\Omega)=\frac{b_0+b_1\Omega+\ldots\ b_{nb}\Omega^{nb}}{a_0+a_1\Omega+\ldots\ a_{na}\Omega^{na}}e^{-j\omega T_d} \qquad (2-36)$$

where Ω is interchangeably *s* or *z*. A range of approaches have been proposed, which are discussed in detail in an excellent survey article by Pintelon *et al.* (1994). Employing the EV model of Figure 2-3, Schoukens *et al.* (1988) derived a maximum likelihood estimator of $H(\Omega)$, which was subsequently developed by Pintelon *et al.* (1992), to arrive at a cost function

$$K(\mathbf{p})=\frac{1}{2}\sum_{k=1}^{F}\frac{|\,e^{-j\omega_k T_d}N(j\omega_k,\mathbf{p})\overline{U}_m(j\omega_k)-D(j\omega_k,\mathbf{p})\overline{Y}_m(j\omega_k)\,|^2}{\sigma_N^2(\omega_k)\,|\,D(j\omega_k,\mathbf{p})\,|^2+\sigma_M^2(\omega_k)\,|\,N(j\omega_k,\mathbf{p})\,|^2-2\Re\{\,\sigma_{MN}(j\omega_k)e^{-j\omega_k T_d}N^*(j\omega_k,\mathbf{p})D(j\omega_k,\mathbf{p})\}} \qquad (2-37)$$

where **p** is the parameter vector

$$\mathbf{p} = [\, a_0 \ldots a_{na} \quad b_0 \ldots b_{nb} \,]^T \tag{2-38}$$

and F is the number of frequencies used in estimation. It can be seen that a nonparametric noise model is employed, in contrast to the time-domain approach, and the noise variances and covariance are required as *a priori* information. These can be estimated from the data before estimation, provided several independent measurements are available, which will be discussed in Chapter 3.

This somewhat daunting cost function can be interpreted as a least squares formulation, with each frequency point weighted by a noise variance term. The larger the noise variance at a given frequency the less that frequency will influence the cost function. The estimator is an MLE under the assumption of complex normal noise, which was discussed earlier.

Under this assumption, double the cost function is the sum of $2F$ squared Gaussian variables and is hence χ^2 distributed with $2F$-n_p degrees of freedom, where n_p is the number of free parameters. Consequently, its expected value will be

$$K_{\min} = F - \frac{n_p}{2} \tag{2-39}$$

The cost function is insensitive to the multiplication of the parameters by a scalar and hence a constraint must be placed on the scaling of the parameter vector. This can be achieved by fixing one nonzero parameter value or by fixing the overall norm of the parameter vector itself. The pure time delay T_d can also be included as a free parameter for estimation, which is an attractive feature of the frequency-domain approach, since its value is not fixed to multiples of the sampling interval. The delay is estimated in seconds for s-domain models and in fractional multiples of the sampling period for z-domain models.

Numerical techniques developed for nonlinear least squares problems, such as Gauss-Newton, Levenberg-Marquardt or singular value decomposition, can be used to find the parameters which minimise the cost function. Since the estimator is an MLE the covariance matrix can once again be approximated by the Cramér-Rao bound, under the

assumption that the true system lies within the range of possible models. The uncertainty on the estimated parameters can then be translated to uncertainties on the estimated poles and zeros, using the analytical technique described by Guillaume *et al.* (1989).

The estimator is now available as part of a Matlab Frequency-domain System Identification Toolbox, written by Kollár (1994) and its derivation and properties are reviewed in a paper by Kollár (1993).

2.3.3 Model Order Selection and Validation

The selection of an appropriate model order and the validation of that model involves techniques which are common to both time- and frequency-domain approaches. The most fundamental is to monitor the behaviour of the cost function as the model order is increased. The cost function should always decrease with increased model order for noisy data and any results to the contrary indicate a poor convergence of the search algorithm.

If good convergence is achieved then a graph of the cost function against model order will usually show a definite knee. This is illustrated in Figure 2-7 for a typical example using the frequency-domain estimator discussed in the previous section. The cost function is plotted against the number of model parameters, along with its expected value and the 95% bound on that value. The benefit of increasing the model order becomes negligible after reaching a certain number of parameters.

Over-modelling can be detected by high parameter uncertainties, which map to high uncertainties on the model poles and zeros. Near-cancelling pole-zero pairs can also appear under these conditions, with uncertainty regions which overlap. Over-modelling can also give rise to models which are not credible, given *a priori* knowledge of the system under test. In the most extreme case, an unstable model can be fitted to data gathered from a system which is known to be stable. Alternatively, poles and zeros can appear at frequencies outside the bandwidth of the estimation data or outside the credible operating dynamics of the system.

The use of periodic test signals allows a high quality FRF to be estimated, which should be free from systematic leakage and aliasing errors if care is taken in the experiment design. This FRF can be used to assess the quality of the estimated model by comparing the model frequency response with the FRF. A poor fit to the frequency data can be an

indication of under-modelling and that the model order should be increased. Alternatively, it may indicate that the frequency data contain significant nonlinear or disturbance effects, which cannot be described by a linear model.

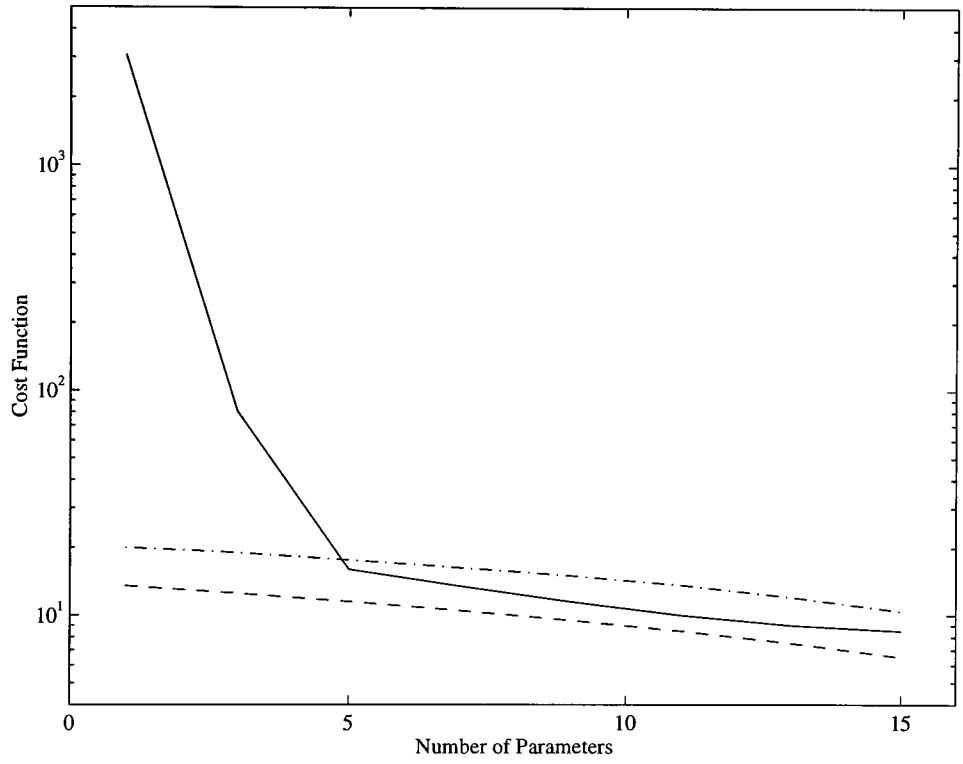

Figure 2-7. Cost function versus number of model parameters (solid), with its expected value (dashed) and the 95% bound on that value (dash-dot).

Once a model structure has been selected in this way it can be further checked by looking at the statistical properties of the residuals. The time-domain residuals defined in equation (2-18) should be approximately independent and normally distributed, with zero mean and a given variance. A visual inspection of the residuals themselves is also useful, since large or untypical values may indicate bad data or *outliers*.

Evaluation of the residuals is not so straightforward in the frequency-domain case, since their calculation requires the estimation of the true input and output Fourier coefficients, which are termed the *nuisance parameters*. It is thus preferable to study the residuals of the cost function itself, also termed the residuals of the noise referred to the output, as discussed for equation (2-35).

The final and decisive test for any estimated model is *cross-validation*, which involves repeating the estimation using a completely new data set and comparing the model parameters. The parameters should be identical, to within the prescribed uncertainty bounds.

2.3.4 Comparison of Models and Approaches

A comparison is made in Table 2-1 of the different properties of s-domain and z-domain models. It can be seen that z-domain models have important advantages for applications such as control system design and system simulation. On the other hand, if a precise description of the system is required, then s-domain models are preferred.

TABLE 2-1

A COMPARISON OF s-DOMAIN AND z-DOMAIN MODELS

Property	s-domain	z-domain
Exact representation of a continuous system	Yes	Only if input is ZOH
Exact representation of the pure time delay	Yes	Only if a multiple of T_s
Simulation	Approximate	Yes
Predictive models	No	Yes
Recursive estimation	No	Yes
Controller design	Limited	Extensive

The choice of model has a direct impact on the type of identification employed, since the estimation of s-domain models is not straightforward in the time domain. Frequency-domain identification allows an s-domain model to be directly identified, whose poles and zeros can then be related to physical parameters of the system. The effects of nonlinearities are also more easily detected in the frequency domain, by the presence of output harmonics in addition to those present in the input signal.

TABLE 2-2

RELATIONSHIP BETWEEN MODELLING GOAL AND CHOICE OF METHOD
(based on Isermann, 1980)

Modelling Goal	Type of Model	Required Accuracy	Identification Method
Fault detection	Linear/nonlinear, parametric	High	On-line parameter estimation
Verification of theoretical models	Linear, continuous-time, nonparametric/ parametric	Medium-high	Off-line, frequency response, parameter estimation
Digital controller design	Linear, discrete-time, parametric	Medium	On-line/off-line, parameter estimation
Adaptive digital control	Linear, discrete-time, parametric	Medium	On-line recursive parameter estimation
Controller tuning	Linear, continuous-time, nonparametric	Low	Off-line, step response

The relationship between the modelling goal and the type of model selected was discussed by Isermann (1980), who compiled a table similar to that shown in Table 2-2. It is clear that if the aim of modelling is to verify theoretical models then continuous-time models are required. In this case, the approach most commonly employed is to estimate discrete-time models in the time-domain and then derive the continuous-time models from them. In most instances this approach is perfectly valid but there can be problems with such a derivation, as will be seen with the gas turbine modelling presented later in this thesis. In such an event, it is preferable to directly estimate continuous-time models in the frequency domain.

2.4 Conclusions

This chapter has served to introduce the modelling concepts and estimation techniques which will be used in this thesis. As such, it is simply a recapitulation of previous results and no original work has been presented.

The important distinction between the true system and the model has been emphasised. A general discussion of linear modelling and time- and frequency-domain identification techniques has been presented. Each of the estimators discussed is an MLE if the disturbing noise is Gaussian. This means that the estimates are asymptotically unbiased, efficient and normal and that the Cramér-Rao lower bound can be used to approximate the covariance matrix of the estimates.

This assumption is only strictly valid if the true system model belongs to the set of models which can be estimated. In reality, this is often not the case and it certainly will not be true for many of the systems studied in this thesis, such as those suffering nonlinear distortions. However, the Cramér-Rao bound can still provide an useful indication of the magnitude of the estimated parameter uncertainties, provided the modelling errors are small.

In the words of Ljung (1987, p. 341): "Analysis pertains to assuming certain properties of the true data-generation mechanism and subsequently calculating the resulting properties of the models. Such calculations turn out to be useful and suggestive, even when the underlying assumptions may not be verifiable."

A brief discussion has also been presented of the relative merits of continuous and discrete models and of time- and frequency-domain estimation. It is concluded that if an exact representation of the continuous system is required, with an arbitrary value of pure time delay, then continuous s-domain models are preferred. Such models can be directly estimated in the frequency domain, using the techniques described in this chapter.

2.5 Appendix - Terms and Concepts in Model Estimation

Any quantities calculated from measurements disturbed by noise will only be estimates of the unknown parameters of the system being modelled. The quality of a given estimate can be assessed by its:

- *Bias* - the systematic error on the value obtained;
- *Variance* - a measure of the uncertainty on the estimate.

The first- and second-order moments of a random variable x are termed its mean and variance and defined as

$$\mu(x) = E\{x\} \qquad (A-1)$$

$$\sigma^2(x) = E\{ (x - E\{x\})^2 \} \qquad (A-2)$$

where $E\{\bullet\}$ denotes the expected value. The standard deviation is the square root of the variance and hence has the same units as the mean value. These quantities must normally be estimated from a finite set of N samples

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad (A-3)$$

$$s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 \qquad (A-4)$$

which are known as the sample mean and sample variance respectively. The variance of the sample mean is given by s^2/N. The sample variance of a complex random variable $z = x + jy$ is given by

$$s_c^2 = \frac{1}{N-1} \sum_{i=1}^{N} (z_i - \bar{z})(z_i - \bar{z})^* \qquad (A-5)$$

If the variable has a complex normal distribution then the real and imaginary parts are independent and normally distributed, with equal variances. The complex variance is

then twice the variance on either the real or imaginary parts.

For a column vector variable $\mathbf{x}$, the covariance matrix is defined as

$$\mathbf{C}(\mathbf{x}) = [\,(\mathbf{x} - E\mathbf{x})(\mathbf{x} - E\mathbf{x})^T\,] \qquad (A-6)$$

where the element c_{ij} of $\mathbf{C}(\mathbf{x})$ is the covariance between elements i and j of $\mathbf{x}$. It can also be expressed in terms of a "mean of the squares" matrix and the square of the means

$$\mathbf{C}(\mathbf{x}) = E\,[\,\mathbf{x}\mathbf{x}^T - E\mathbf{x}\cdot\mathbf{x}^T - \mathbf{x}E\mathbf{x}^T + E\mathbf{x}\cdot E\mathbf{x}^T] = E\,[\,\mathbf{x}\mathbf{x}^T\,] - E\mathbf{x}E\mathbf{x}^T \qquad (A-7)$$

There are a number of key definitions in the classification of estimators:

Unbiased: an estimator is unbiased if the mathematical expectation of the estimates equals the true parameter values, $\mathbf{p}_0$

$$E\{\hat{\mathbf{p}}\} = \mathbf{p}_0 \qquad (A-8)$$

Asymptotically unbiased: if the expected value converges to the true value as the number of measurements N increases

$$\lim_{N\to\infty} E\{\hat{\mathbf{p}}_N\} = \mathbf{p}_0 \qquad (A-9)$$

where N denotes the number of data points, which can be interchangeably the number of time-domain samples or the number of frequencies.

Consistent: an estimator is consistent if $\hat{\mathbf{p}}_N$ converges in some sense to $\mathbf{p}_0$ as N approaches infinity. Several convergence concepts have been defined, which include convergence:

— in mean square $$\lim_{N\to\infty} E\{\,(\hat{\mathbf{p}}_N - \mathbf{p}_0)^2\,\} = 0 \qquad (A-10)$$

— with probability one $$\lim_{N\to\infty} P\{\,\hat{\mathbf{p}}_N = \mathbf{p}_0\,\} = 1 \qquad (A-11)$$

— in probability $$\lim_{N\to\infty} P\{\,|\,\hat{\mathbf{p}}_N - \mathbf{p}_0\,| > \delta\,\} = 0 \qquad \forall \quad \delta > 0 \qquad (A-12)$$

where the probability of the bias being greater than any nonzero value δ is zero as $N \to \infty$.

Convergence in mean square implies that the estimator is asymptotically unbiased. Convergence in mean square and convergence with probability 1 are affinal, but not identical, properties. Convergence in mean square and with probability one both imply convergence in probability but not vice versa. An estimator which exhibits convergence with probability one is termed *strongly consistent*, while an estimator which only exhibits convergence in probability is termed *weakly consistent.*

Efficient: an unbiased estimator E_1 is efficient if $\mathbf{C}_1 < \mathbf{C}_2$ for all unbiased estimators E_2, where $\mathbf{C}_1$ and $\mathbf{C}_2$ are the parameter covariance matrices, which relate the true parameter values $\mathbf{p}_0$ and the estimated values $\hat{\mathbf{p}}_N$

$$\mathbf{C}(\hat{\mathbf{p}}_N) = E\{\,[\hat{\mathbf{p}}_N - \mathbf{p}_0][\hat{\mathbf{p}}_N - \mathbf{p}_0]^T\,\} \qquad (A-13)$$

The variances of the individual elements of $\hat{\mathbf{p}}_N$ make up the principal diagonal of $\mathbf{C}(\hat{\mathbf{p}}_N)$. A desirable property of any estimator is that the variance, or uncertainty, of the estimates is as small as possible. However, there exists a lower bound on the covariance matrix of an unbiased estimator, termed the *Cramér-Rao bound*, which is expressed as

$$\mathbf{C}(\hat{\mathbf{p}}_N) \geq \mathbf{F}^{-1} \qquad (A-14)$$

where $\mathbf{F}$ is the *Fisher information matrix*, a measure of the amount of information present, in relation to the parameters. It is possible to construct an estimator with a covariance matrix lower than the Cramér-Rao bound but it will be biased. The information matrix applies to any unbiased estimator of the parameters $\mathbf{p}$ using the measurements y

$$\mathbf{F} = E\left\{\left(\frac{\partial \ln p(y \mid \mathbf{p})}{\partial \mathbf{p}}\right)\left(\frac{\partial \ln p(y \mid \mathbf{p})}{\partial \mathbf{p}}\right)^T\right\}\Big|_{\mathbf{p}=\mathbf{p}_0} = -E\left\{\left(\frac{\partial^2 \ln p(y \mid \mathbf{p})}{\partial \mathbf{p}^2}\right)\right\}\Big|_{\mathbf{p}=\mathbf{p}_0} \qquad (A-15)$$

where $p(y \mid \mathbf{p})$ is the joint probability density function of the output in relation to the parameter vector and the matrix of second derivatives is known as the Hessian. If $\mathbf{p}$ is a d-dimensional column vector then the first derivatives will also form a d-dimensional column vector and the Hessian will be a $d \times d$ matrix.

Normal: it is highly desirable that the parameter estimates are normally distributed, since a normally distributed variable is completely defined by its first- and second-order

moments and this allows statistical bounds to be calculated in a straightforward manner.

Robust: if some properties of the estimator are still valid when not all the *a priori* assumptions are met. Most commonly this relates to a violation of the assumptions regarding the noise. If the estimator can still produce a consistent estimate then it will converge to the true value as $N \rightarrow \infty$.

Maximum Likelihood Estimator (MLE): the MLE is a strongly consistent and asymptotically unbiased estimator if the true system belongs to the estimator model set. In the presence of *independent, identically distributed* (iid) measurements it is also asymptotically efficient, since it approaches the Cramér-Rao lower bound asymptotically. If there is an estimator which reaches this lower bound it will be an MLE. The parameter estimates are also asymptotically normal.

The primary sources for the summary presented in this Appendix were:

Åström, K.J. (1980). "Maximum likelihood and prediction error methods", *Automatica,* vol. 16, pp. 551-574.

Norton, J.P. (1986). *An Introduction to Identification*, Academic Press, London.

Ljung, L. (1987). *System Identification - theory for the user,* Prentice Hall, Englewood Cliffs.

Chapter III

Experiment Design

Abstract — *A discussion is presented of the key issues in experiment design. Particular attention is paid to the selection of the input excitation and the use of periodic signals is advocated. The properties of single sines, multisines and maximum length binary sequences are discussed and data analysis techniques which can be employed with periodic signals are outlined. The relative test times for single sine and multisine signals are then compared. A frequency-domain approach is proposed for binary signal design.*

3.1 Introduction

The data gathering experiment is a key step in any identification procedure, since the quality of the final model will directly depend on the quality of the data from which it is estimated. Data gathering can also be the most time consuming and expensive aspect of modelling and hence the design of the experiment must be given careful consideration. It was summarised by Isermann (1980) as the choice of:

- Input signals, their amplitude and spectra;
- Sampling frequency;
- Signal filtering;
- Equipment for signal generation and measurement;
- Measurement time;
- Off-line or on-line estimation;
- Open-loop or closed-loop operation.

The design of appropriate input signals is the main topic of this chapter and indeed this thesis. Only the influence of stochastic errors will be considered in this chapter, since the effects of nonlinear distortions are dealt with in Chapters 4 and 5. The choice of sampling frequency and signal filtering will be seen to depend on the basic assumptions associated with continuous or discrete modelling, as will the test equipment required. The measurement time will be shown to be a function of the required model accuracy, although this can often be limited by financial or operational constraints. Only off-line estimation will be considered in this thesis, since frequency domain identification requires the batch processing of data using the FFT. It is also assumed that the system under test is operating in open loop.

3.2 Basic Assumptions

The basic assumption for the identification of a continuous s-domain model is that the data have been sampled without aliasing. This implies that the input and output signals have been band-limited before sampling and sampled at a frequency f_s, which is more then twice the highest frequency f_{max}. This will be termed the *band-limited* (BL) assumption.

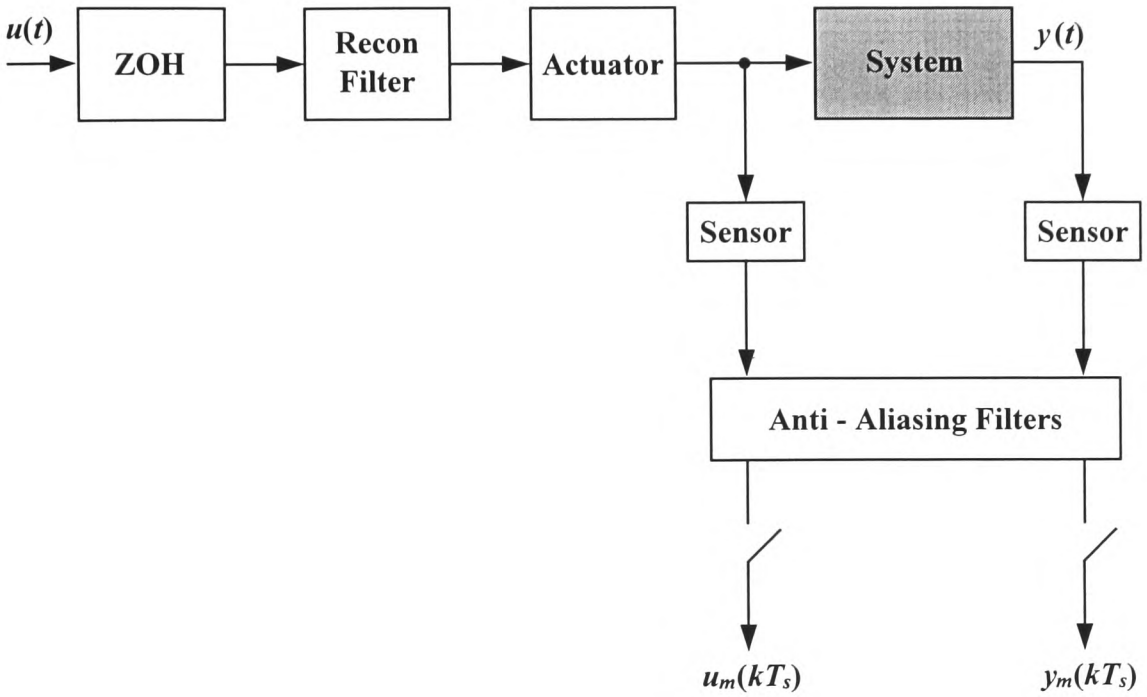

Figure 3-1. Experimental set-up under BL assumption.

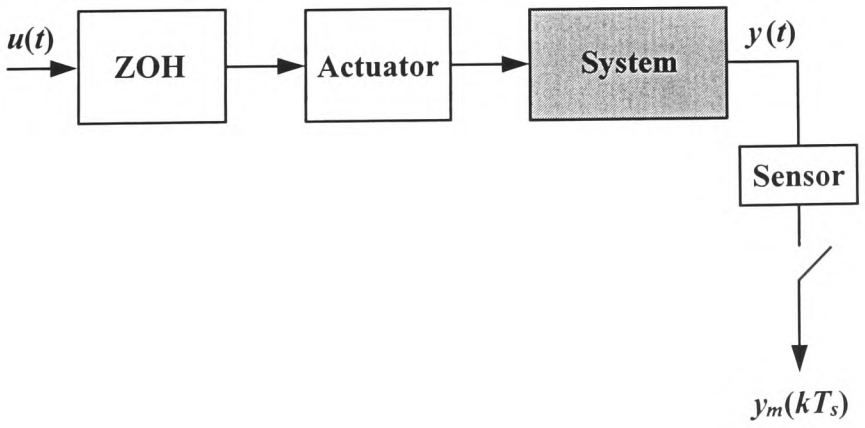

Figure 3-2. Experimental set-up under ZOH assumption.

The data must also be gathered in dynamic steady state, which implies that periodic excitation signals are used and that all transient effects have died away. The signals must be sampled over an integer number of periods, in order to avoid leakage errors in the FFT, using a sampling frequency which is a multiple of the signal fundamental frequency.

A suitable test set-up is shown in Figure 3-1, where an analogue *reconstruction* filter is used to filter out the higher frequency components at the output of the ZOH. These components are centred at multiples of the reconstruction frequency and are often referred to as *frequency butterflies* because of their shape. They are the result of the frequency-domain convolution of the sampled signal, which is periodic in the frequency domain, with the transfer function of the ZOH. A reconstruction filter should not be used with binary input signals.

Analogue *anti-aliasing* (AA) filters are used to band-limit the signals before sampling and their cut-off frequencies must be carefully chosen, to ensure that they provide sufficient attenuation at the Nyquist frequency but do not filter out important information at the test frequencies. An useful guideline is to set the sampling frequency at least ten times the maximum frequency of the signal and then to set the AA cut-off frequency at $2.5f_{max}$. This ensures that the cut-off frequency is sufficiently distant from f_{max}, while providing a frequency span of at least one octave between the cut-off and Nyquist frequencies.

The measurement channels should have identical characteristics, in order that their influence on the input and output signals is the same. This can be difficult to achieve in practice and a relative calibration may need to be performed in order to equalise their effects (Schoukens *et al.*, 1994a).

A discrete model is only an exact representation of a continuous process if the input signal is piece-wise constant between sampling instants, which was discussed in Chapter 2 and termed the ZOH assumption. This point is particularly important if a physical interpretation is to be made of the estimated model and is discussed in detail by Andersson *et al.* (1994) and Schoukens *et al.* (1994a). The assumption has important implications for the conditions under which a discrete identification experiment is performed. The signal applied to the system should be generated by a ZOH, which means that no reconstruction filter should be used. In addition, no AA filters should be used in measurement, otherwise important information will be lost. If the ZOH is assumed to be perfect then there is no need to sample the input signal, which simplifies the measurement set-up, as shown in Figure 3-2.

It is clear from Figure 3-2 that a model fitted between $u(t)$ and $y(t)$ will include the ZOH, actuator and sensor, as well as the system under test. This is not a problem if the overall objective is control but if an accurate model of the system is required then their influence should, as far as possible, be excluded. It is then desirable to measure the input signal after the actuator, since this represents the true input to the system. If the actuator band-limits the input signal the ZOH assumption will be violated, which will introduce errors into the estimated discrete model (Schoukens *et al.*, 1994a).

The preferred approach will depend on the intended application of the model but if accuracy and a physical interpretation of the model parameters are primary considerations then the BL assumption should be selected.

Under the BL assumption the sampling frequency must be at least twice the highest frequency in the signal, though in practice, a value at least ten times the highest frequency is often selected. Increasing the sampling frequency has no adverse effects on frequency-domain modelling and, in the presence of white noise, it has the positive effect of improving the SNR, by spreading the noise power across a wider frequency band.

The selection of the sampling frequency is far more critical for time-domain modelling, since it directly affects the quality of the estimates (Goodwin and Payne, 1976). If the sampling frequency is too low then important information about the system will be lost and an inaccurate model will result. If the sampling rate is too fast then numerical problems can result, due to the clustering of the discrete poles and zeros around the point $z = 1$. Isermann (1980) suggested that the limits of the sampling interval should be set at

$$\frac{T_{95}}{T_s} \approx 5 \ldots 15 \qquad (3-1)$$

where T_{95} is the 95% transient settling time and T_s is the sampling interval. A sampling frequency of around ten times the highest break-point frequency has been recommended by Norton (1986, p. 12) and by Ljung (1987, pp. 378-386), which corresponds to the lower limit of the ratio in equation (3-1). Ljung also pointed out that the estimation results will be far worse if the sampling frequency is too low, rather than too high.

3.3 Input Signals

A wide variety of signals may be used to excite a system during an identification experiment. Classical techniques make use of transient signals such as impulse and step inputs. These signals are susceptible to noise and attempts to overcome this may drive the system out of its normal operating range. An alternative is to use a random signal such as Gaussian noise but once again the SNR may be poor and then power spectrum averaging will be necessary in order to obtain accurate FRF estimates.

Periodic signals are an attractive alternative to both transient and random signals and the simplest of these are monotonic sine waves. The use of periodic binary sequences has been widespread since the late 1960's, since they are easy to generate and have good auto-correlation properties. Improvements in the speed and cost of computer memory have now made the generation of non-binary multifrequency signals, termed *multisines*, a practical option.

A minimum requirement for any test signal is that it excites all the modes of the system, such that the experiment is said to be informative. An open-loop experiment is informative if the input signal is *persistently exciting* (PE). A signal is PE of order n if its spectral density is non-zero for at least n distinct points in the interval $(-\pi,\pi)$, which can be generated by the sum of $(n+1)/2$ sinusoids. A signal which is PE of order n cannot be filtered to zero by a moving average filter of order $(n-1)$. Van den Bos (1970, p. 26) showed that a signal of at least $(n+1)/2$ sinusoids is required to identify an s-domain model with n parameters and the same result was derived for z-domain models by Ljung (1987, pp. 362-364).

The use of signals which are on the limit of persistence of excitation, i.e. using a signal with $(n+1)/2$ harmonics to identify a model with n parameters, is not recommended. Such an approach leaves no room for an increase in the model order, or the detection of model errors through the comparison of the model frequency response with the estimated FRF.

3.3.1 Periodic Signals

Periodic test signals are a requirement for frequency-domain identification if systematic errors are to be avoided and they also offer a number of advantages for time-domain identification. These were discussed by Schoukens *et al.* (1994b) and the most important can be summarised as follows.

Improvement in the Signal to Noise Ratio. The crest factors of periodic signals are generally much better than those of random or transient signals, which allows more power to be injected for a given time-domain amplitude. Since the signals have discrete spectra it is also possible to select only those frequency lines at which power is injected and discard the noise lines.

Model Validation. Periodic signals can be transformed to the frequency domain without leakage errors and if the SNR is good a high quality estimate of the FRF can be obtained. This allows the validation of both continuous and discrete parametric models by comparing the frequency response of the model with the estimated FRF.

Combination of Data Sets. Data sets covering different frequency ranges, gathered at different sampling rates, can be transformed to the frequency domain and combined into one global data set. Such combination is not possible with nonperiodic time-domain data.

Improvement of Finite Sample Behaviour. Only those frequencies at which power is injected need be used in estimation. For time-domain estimation this can be achieved by transforming the signals to the frequency domain, setting the noise contributions at all other frequencies to zero and then calculating the inverse transform. This has been shown to have a large impact on the convergence region of both time- and frequency-domain estimators.

A further advantage can be added to this list, which will prove to be particularly useful in the time-domain modelling of gas turbine dynamics, discussed in Chapter 9.

Data Resampling. The minimum sampling interval for time-domain modelling is usually set by the sampling frequency used in data gathering. The use of periodic signals allows the sampling interval to be decreased in a straightforward manner, by calculating the FFT of the signal and then taking the inverse Fourier transform at a faster sampling rate.

3.3.2 Signal Quality Measures

In discussing signal quality it is important to stress from the outset that the best signal for the estimation of the nonparametric FRF may not be the best for parametric estimation. The worst case accuracy of the FRF will depend on the power in the weakest spectral components, or the points at which the disturbing noise is strongest. This is not the case for parametric estimates, since their quality depends on the total information in the data

set. But since an FRF estimate will still be required to validate the parametric model, the importance of a high quality FRF estimate remains. For this reason, the quality of signals in relation to FRF estimation will now be discussed in detail.

Signal quality can be assessed by a number of different criteria and by far the most widely used is the *crest factor* (CF), defined as

$$\mathrm{CF} = \frac{\max\{|u(t)|\}}{\sqrt{\overline{u^2(t)}}} = \frac{u_{\max}}{u_{\mathrm{rms}}} \qquad (3-2)$$

which is a measure of the amplitude compression of the signal in the time domain, for a given *root mean square* (RMS) value. The main drawback of the CF as a comparative measure is that no account is taken of the spectral amplitude of the signals. This means that a signal can have a good CF but low power at certain frequencies, resulting in low quality FRF estimates at these points.

An alternative measure, termed the *time factor* (TF), was proposed by Schoukens *et al.* (1993a) to assess the quality of a signals for measurement of the FRF. It was derived under the assumption that the input and output noises are uncorrelated and is based on the uncertainty of the FRF estimates defined in equation (2-34). The number of averages M, necessary to achieve an uncertainty of σ^2_{Hmin} can be expressed as

$$M = \max\left\{ \frac{2}{\sigma^2_{Hmin}} \left(\frac{|H(j\omega_k)|^2 \sigma^2_M(\omega_k) + \sigma^2_N(\omega_k)}{|U(j\omega_k)|^2} \right) \right\} \qquad k = 1, 2, \ldots, F \qquad (3-3)$$

The equation can be simplified by assuming that the noise referred to the output, $|H(j\omega_k)|^2 \sigma^2_M(\omega_k) + \sigma^2_N(\omega_k)$ is a constant, independent of the chosen input signal. The term σ^2_{Hmin} is similarly an arbitrary constant. Since the measurement time is proportional to the number of averages then the measurement time per frequency point will be proportional to M/F.

$$\frac{M}{F} = \max\left\{ \frac{1}{|U(j\omega_k)|^2} \frac{1}{F} \right\} \qquad k = 1, 2, \ldots, F \qquad (3-4)$$

The magnitude of the terms on the right-hand side of equation (3-4) can be reduced simply by increasing the frequency amplitudes. A constraint must therefore by placed on the

time-domain amplitude of the signal, which leads to the final definition of the time factor

$$\mathrm{TF} = \max\left\{ \frac{u_{\max}^2}{|U(j\omega_k)|^2} \frac{1}{F} \right\} \qquad k = 1, 2, \ldots, F \tag{3-5}$$

where $U(j\omega_k)$ is the single sided spectrum and F is the number of frequency measurements. The measurement time required to achieve a specific minimum accuracy with different test signals is proportional to their relative time factors. The utility of the CF and TF as comparative measures for single sine and multisine signals will be discussed in Section 3.5.

3.3.3 Single Sines

Single sine testing involves injecting a pure sinusoid at a range of different frequencies and estimating the FRF at each frequency point.

$$u(t) = A \cos(\omega t) \tag{3-6}$$

These are the most commonly used periodic signals and they have a number of attractive features, including ease of generation and the straightforward detection of nonlinear effects by visual inspection of the time and frequency records. Their main drawback is that a settling time must be allowed after each change of frequency, in order to allow the system to reach dynamic steady state. Since the total signal power is concentrated at a single frequency the SNR obtained with single sine testing will be the highest possible, for a non-binary signal, at a given maximum input amplitude. The total test time for a set of F equally spaced measurements, at integer multiples of a fundamental frequency f_0, will be

$$T_{\text{test}} = M \frac{1}{f_0} F + T_{\text{set}} F \tag{3-7}$$

where T_{set} is the settling time allowed after each change of frequency before commencing measurements. This expression assumes that each sine is measured for the same time period in order to ensure equal accuracy at each frequency. The CF of a single sine will always be $\sqrt{2}$ and its TF will be 1.

3.3.4 Multisines

These are an arbitrary ensemble of harmonically related cosines

$$u(t) = \sum_{k=1}^{F} a(k) \cos\left(i(k)\, \omega_0\, t + \phi(k) \right) \qquad (3-8)$$

with **a** a vector of amplitudes, **i** a vector of harmonic numbers, ω_0 the signal fundamental, Φ a vector of phases and F the number of cosines in the signal. The relative phases of the harmonics must be carefully selected to minimise the signal CF. This problem has been addressed by many authors, and while no analytical solution has yet been found a number of approaches yield significantly improved values. The first and most empirical method was proposed by Schroeder (1970), who observed that phase modulated signals had low peak factors and derived a formula

$$\phi(1) = 0 \qquad \phi(k) = 2\pi \sum_{l=1}^{k-1} (k-l)\, p_l \qquad k = 2, 3, \ldots, F \qquad (3-9)$$

where p_l is the power of the lth harmonic and $\sum_{l=1}^{F} p_l = 1$.

For signals with a flat amplitude spectrum the formula can be reduced to

$$\phi(1) = 0 \qquad \phi(k) = \pi \frac{k^2}{F} \qquad k = 2, 3, \ldots, F \qquad (3-10)$$

Multisines with Schroeder phases were used in the testing and modelling of a number of different industrial systems by Flower and Forge (1981). The Schroeder formulae result in low CF for signals with dense, evenly-spaced harmonics but do not work well on signals with sparse spectra, such as log-tones (Van der Ouderaa and Renneboog, 1988). Other methods based on explicit formulae have been proposed by Shapiro and Rudin and by Newman, which were discussed by Boyd (1986) and Van der Ouderaa *et al.* (1987). Each method generates signals with CF similar to those obtained using Schroeder phases.

Rees (1986) and Rees and Jones (1991) considered CF minimisation as the minimisation of a non-linear function of F independent phases and used a quasi-Newton algorithm to search for the minimum. This approach yields lower CF than the Schroeder phases but

requires a considerable amount of processing time. The result is also very sensitive to the phases used as starting values, since the function has a large number of local minima. The derivative of the CF with respect to the phases is also discontinuous (Van der Ouderaa, 1988, pp. 63-65), which means that any gradient-based approach will not find the true minimum of the CF. A related approach using simulated annealing was presented by Van den Bos (1993), who was reporting on work conducted by Lenstra. The use of simulated annealing reduces the likelihood of the search becoming trapped in a local minimum but the method is still extremely time consuming.

Van den Bos (1987) proposed an iterative method which attempts to make the multisine signal maximally similar to a binary signal by adjusting its phase angles, based on the observation that binary signals possess the minimum possible CF. This is achieved by first constructing a multisine signal with the required amplitude spectrum and random phases. The values of this signal at discrete points are used to construct an equivalent binary signal, with positive and negative values mapping to +1 and -1. Zero values are arbitrarily mapped to either +1 or -1. The FFT of this binary signal yields the amplitude and phase of the required harmonics and these are then used to reconstruct a multisine. The process is repeated until the multisine and binary signals have equal phases at the harmonics of interest.

A closely related approach was proposed by Van der Ouderaa *et al.* (1988) but instead of constructing a binary signal with each iteration the multisine is clipped to a certain percentage of its initial peak amplitude. Both the Van den Bos and the Van der Ouderaa methods are classed as clipping techniques in Table 3-1.

More recently, Guillaume *et al.* (1991) proposed a method derived from the observation that the CF can be expressed as

$$\mathrm{CF} = \frac{l_\infty(u)}{l_2(u)} \qquad (3-11)$$

where $l_\infty(u)$ is the Chebyshev norm of $u(t)$ and $l_2(u)$ its RMS value. Since l_2 is independent of Φ the problem reduces to the minimisation of the l_∞ norm with respect to the phases. Since this is non-differentiable, the l_{2p} norm is minimised in its place, using a

Levenberg-Marquardt algorithm. The norm is initially minimised with a small value of p, which is then repeatedly incremented, with the phases of the previous minimisation being used as starting values for each new step.

Simulations have shown that it produces the lowest CF achieved to date for band-limited multisines and it is also fast enough for practical application. Another important advantage of this approach is that it can be readily used to minimise the CF's of signals at the input and the output of a linear system. This will be termed the $L\infty$ method and a comparison of the four main classes of CF minimisation techniques is made in Table 3-1.

TABLE 3-1

COMPARISON OF CREST FACTOR MINIMISATION METHODS

Properties	Formulae	Search	Clipping	$L\infty$
Simplicity	Good	Fair	Poor	Poor
Speed	Good	Poor	Fair	Fair
Arbitrary Spectrum	Poor	Good	Good	Good
Approaches Optimum	Poor	Fair	Good	Good
Input-Output Optimisation	No	No	Limited	Yes
Analytical Basis	Poor	Fair	Good	Good

If the fundamental and all the harmonics over a given bandwidth are included in the signal it is termed a *consecutive multisine.* The great flexibility of multisines is that their amplitude spectrum can be set to any arbitrary value. A signal composed of only those harmonics which are odd multiples of the fundamental will be immune to even-order nonlinear distortions, which will be discussed in greater detail in Chapter 4. Such signals will be termed *odd multisines.* The variation of the CF of consecutive and odd multisines with the number of harmonics is shown in Figures 3-3 and 3-4, for the Schroeder and $L\infty$ approaches.

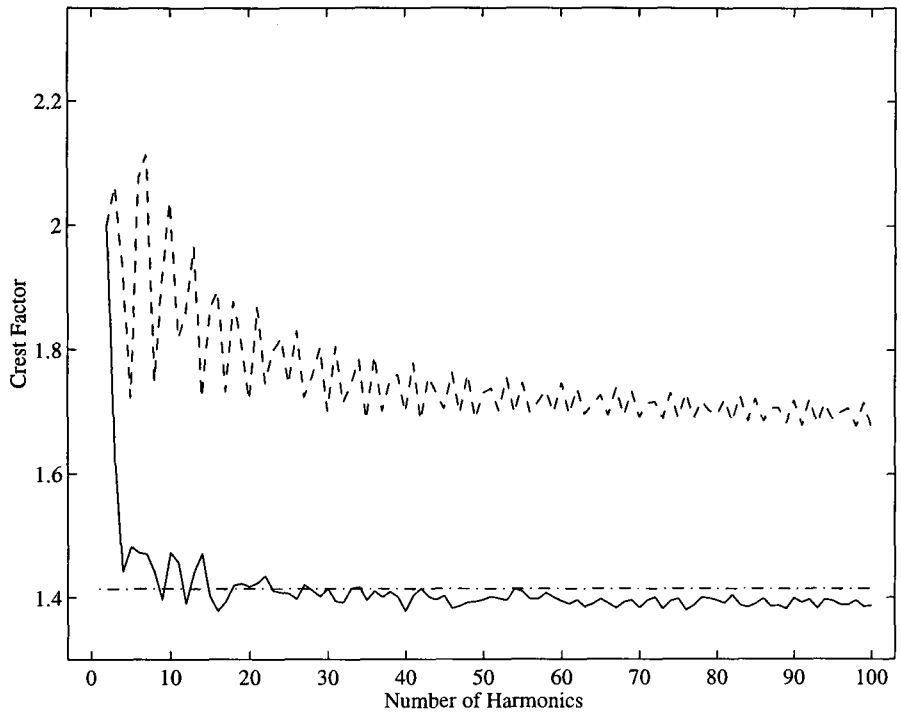

Figure 3-3. Variation of CF for a consecutive multisine, with $L\infty$ phases (solid) and Schroeder phases (dashed). The CF of a single sine is shown as dash-dot.

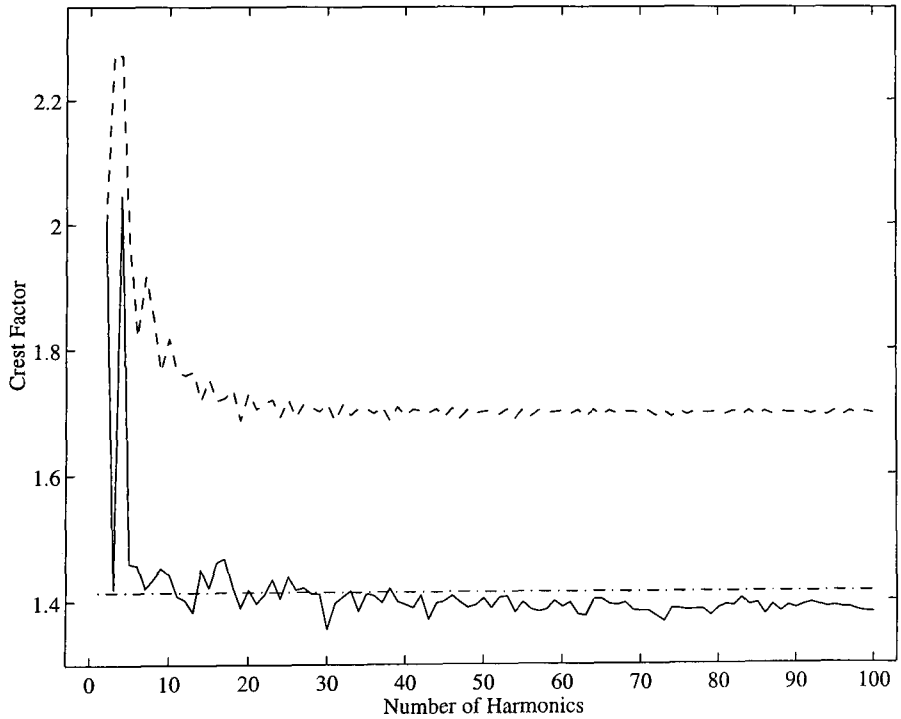

Figure 3-4. Variation of CF for an odd multisine, with $L\infty$ phases (solid) and Schroeder phases (dashed). The CF of a single sine is shown as dash-dot.

It can be seen that both consecutive and odd signals are amenable to CF minimisation and that the $L\infty$ approach gives a significant reduction in CF compared to the Schroeder formula. Indeed, once the number of harmonics has reached 30 or greater, the multisines compressed using the $L\infty$ method have a CF equal to or a little less than that of a single sine.

The total test time for a multisine signal will be

$$T_{test} = M \frac{1}{f_0} + T_{set} \qquad (3-12)$$

The dominant term in this equation is that associated with the number of averages M. The best case multisine CF will be around $\sqrt{2}$ and the best case TF will be around 1.

3.3.5 Binary Signals

These can adopt only one of two values, corresponding to logic levels 1 and 0 and usually assigned the values $+V$ and $-V$. Of the many binary signals which have been described by far the most popular have been *maximum length binary sequences* (MLBS). These are easily generated using shift registers of different lengths with appropriate feedback. The signals are periodic, of sequence length

$$N = 2^n - 1 \qquad (3-13)$$

where n is the number of stages in the shift register. The spectrum of the signal decays as a function of frequency with a sinc x envelope. If the signal has an amplitude of $\pm V$, its single sided amplitude spectrum will be

$$A(\omega_k) = 2V \frac{\sqrt{N+1}}{N} \left| \frac{\sin(k\pi/N)}{(k\pi/N)} \right| \qquad k = 1, 2, 3 \ldots \infty \qquad (3-14)$$

The frequency of the harmonics is dictated by the bit period Δt and N:

$$f_k = \frac{k}{N\,\Delta t} \qquad (3-15)$$

and the half-power bandwidth of the signal occurs at approximately 0.443/Δt Hz (Godfrey, 1993b). It is thus possible to adjust N and Δt to alter the frequency range of the signal. The major disadvantage is that however well it is designed a significant amount of signal power will be spread at unwanted frequencies.

The time-domain design criteria employed with MLBS signals are based on their use with correlation techniques to measure the impulse response of a system. The autocorrelation of an MLBS sequence is approximately impulsive and this can be used to deconvolve the Wiener-Hopf equation and estimate the system impulse response (Godfrey, 1980). In order that the system impulse response does not vary significantly over the width of the autocorrelation pulse it is required that

$$\Delta t \ll \tau \qquad (3-16)$$

where τ is the time constant of the system, and additionally that

$$N\,\Delta t > T_{set} \qquad (3-17)$$

This corresponds to a frequency-domain criterion that the amplitude spectrum of the MLBS signal does not vary significantly over the bandwidth of the system.

A binary signal composed of only odd harmonics can by generated by the modulo-two addition of the bit sequence of an MLBS signal with a sequence of alternating ones and zeros (Godfrey, 1993a). If the addition is carried out across two periods of the original MLBS signal an *inverse repeat* (IR) sequence will be generated, with the required odd harmonic properties.

The total test time for an MLBS signal will be:

$$T_{test} = M\,N\,\Delta t + T_{set} \qquad (3-18)$$

In a similar fashion to the multisine signal, the dominant term is associated with the number of averages. The CF of an MLBS sequence is always 1 and the best case TF is around 1.

Van den Bos (1967) addressed the problem that significant amounts of power in an MLBS signal are at unwanted frequencies, by proposing a random search method to find the binary sequence which best concentrated the signal power in the harmonics of interest. This was later refined to a faster iterative technique which swaps between the time and frequency domains, to produce a signal that approximates, in a least squares sense, the desired spectrum (Van den Bos and Krol, 1979).

A further improvement can be achieved by sequentially swapping the signs of two consecutive bits and retaining this combination if the power spectrum is improved (Paehlike and Rake, 1979). Such signals, termed *discrete interval binary sequences*, give a considerable improvement in the percentage of signal power at the frequencies of interest but significant power is still placed at other frequencies and the power at the frequencies of interest can vary substantially.

3.3.6 Optimal Signals

Optimal experimental design involves selecting an input signal which minimises some measure of the uncertainty on the estimates, for a given system and noise characteristics. For estimation of the FRF this can be defined as maximising the minimum accuracy, for a fixed measurement time and maximum peak value of the signal. No global solution has yet been found to this problem but Schoukens *et al.* (1993a) have proposed a two step approach, which first involves setting the spectral amplitude vector proportional to the standard deviation of the noise, referred to the output

$$a(k) \propto \sqrt{|H(j\omega_k)|^2 \sigma_M^2(\omega_k) + \sigma_N^2(\omega_k)} \qquad (3-19)$$

The CF of the resulting signal is then minimised in the second step. Substitution of equation (3-19) into (2-34) shows that the resulting FRF estimates will have a constant absolute uncertainty.

Considerable attention has been paid to optimal signal design for parametric time-domain estimation (Zarrop (1979), Mehra (1981), Walter and Pronzato (1997)) and it is possible to define a number of optimality criteria, based on scalar functions of the covariance matrix **C**. The most widely used has been the minimisation of the determinant of **C**,

known as D-optimality. In practice, it is most common to maximise the determinant of the Fisher information matrix **F**, since the Cramér-Rao lower bound is easier to calculate than **C**.

The problem immediately arises that **F** is a function of the unknown parameters $\mathbf{p}_0$ and the values expressed in **F** only have a meaning if the estimator is unbiased or at least consistent. If the estimator is known to be consistent, this can be overcome by replacing the unknown parameters $\mathbf{p}_0$ with the best estimates of the parameters obtained from a pilot experiment. Optimal frequency domain design results in signals with a limited number of sinusoids and Goodwin and Payne (1977, p. 151) have shown that it can be achieved with no more than $[n(n+1)/2 + 1]$ sinusoids, where n is the number of unknown parameters.

Schoukens *et al.* (1993a) developed a frequency-domain method based on the *response dispersion function*, which for an input power spectrum

$$P(j\omega) = (\,|\,U(j\omega_1)\,|^2 \ldots |\,U(j\omega_F)\,|^2\,) \quad \text{with} \quad \sum_{k=1}^{F} |\,U(j\omega_k)\,|^2 = 1 \tag{3-20}$$

is defined as

$$v(j\omega) = \text{trace}\{\,[\mathbf{F}(P)^{-1}\,\mathbf{f}(j\omega)]\,\} \tag{3-21}$$

where $\mathbf{F}(P)$ is the information matrix resulting from the power spectrum $P(j\omega)$ and $\mathbf{f}(j\omega)$ the information matrix corresponding to a single frequency $j\omega$, with unity power. The dispersion function can be thought of as the ratio of the variance of the model frequency response to the noise power referred to the system output, at a given frequency $j\omega$. The maximum of $v(j\omega)$ over the signal frequency range will be larger than or equal to the number of model parameters.

The method employed by Schoukens *et al.* restricts the possible spectral components of the optimal signal to discrete harmonically-related frequencies. An initial flat amplitude spectrum is constructed and the dispersion function calculated for every frequency in the signal. Power is then redistributed amongst these frequencies in proportion to the corresponding values of the dispersion function. This procedure is repeated until there is little improvement in the determinant of **F**.

The optimal signals for low-order systems have a small number of harmonics concentrated in the bandwidth of the system, which means that their CF's tend to be high. It is thus preferable to select a sub-optimal solution, by stopping the optimisation after a few iterations. At this point the most important increase in information will have occurred and there will still be sufficient harmonics present to achieve a low CF and also to allow the detection of any modelling errors.

The use of optimal excitations is strictly limited by the complexity of the algorithms involved and the need for an *a priori* model of the system. In addition, if the real system differs from the model used in optimisation the resulting data can be of poorer quality than those obtained using well designed non-optimal signals. Since the benefit is at best a two to one improvement in parameter variances, the additional effort involved is highly questionable and the use of optimal signals is generally not recommended. However, if some information about the system is available *a priori* then an optimal signal can be designed in order to provide a general guide to the frequency bands in which the input signal power should be placed.

3.4 Analysing Periodic Data

The use of periodic signals greatly simplifies the analysis of the measured data. The presence of errors such as poor sampling synchronisation, process drift and measurement outliers can be detected from the time-domain records. Providing the data are sampled across complete periods, at a sampling frequency which is an integer multiple of the fundamental frequency, then no leakage errors will be introduced by the FFT. Statistical information about the frequency data can then be obtained by applying a period by period analysis. This allows a straightforward assessment of the levels of noise and nonlinearities from the frequency-domain records. The following outline owes much to the excellent paper by Kollár *et al.* (1994).

3.4.1 Synchronisation

It is important that the signal reconstruction and sampling clocks are synchronised when conducting measurements for frequency-domain identification, since a deviation in either clock will result in incomplete signal periods being sampled. The quality of the synchronisation can be checked by examining the circular covariance of the measured signals

$$c(nT_s) = T_s \sum_{m=0}^{N-1} u(m)\, u(n \ominus m) \qquad n = 0, 1, \ldots, N-1 \qquad (3-22)$$

where T_s is the sampling period, N the number of samples and $u(n \ominus m)$ corresponds to a circular shift of m samples to the right. This can be calculated via the inverse FFT of the signal power spectra, with the dc component suppressed (Kollár *et al.*, 1994). The circular covariance of periodic data will have peaks at intervals corresponding to the signal period. Any problems with the sampling synchronisation will cause these peaks to appear at irregular time intervals.

3.4.2 Drift and Repeatability

The presence of a drift in the operating point during the experiment can be checked by calculating the mean values of each period of the measured input and output. The presence of significant drift will affect the frequency-domain estimation, since low-frequency transients will be introduced which will corrupt the Fourier coefficients of the measured data. Time-domain estimation of discrete models will also be adversely affected.

Application of periodic signals also allows the repeatability of the system response to be checked, by plotting the outputs period by period, with their mean values removed. This also allows the visual detection of outliers which can then be eliminated by substituting values calculated by interpolation between the previous and next data points.

3.4.3 Noise Analysis

A first indication of the noise levels present in the data can be obtained by transforming the complete data lengths into the frequency-domain using the FFT. Since the frequency resolution of the FFT is inversely proportional to the total time of the data record ($\Delta f = 1/T_{test}$), this will give the greatest resolution achievable with the data. It is then possible to calculate the total power in the signal P_{tot}, and the power at the excited harmonics P_{test}. An initial estimate of the SNR can then be obtained by treating all the power at nonexcited frequencies as noise

$$\mathrm{SNR}_{td} = 10 \log_{10}\left(\frac{P_{test}}{P_{tot} - P_{test}}\right) \qquad (3-23)$$

This is a measure of the total excitation power over the total remaining power and is hence a calculation of the raw time-domain SNR, which can be improved by further processing.

The next step in the noise analysis is to estimate the sample means, variances and covariance of the input and output Fourier coefficients, which are needed both to estimate the uncertainty on the FRF estimates and as *a priori* information for the parametric frequency-domain estimator. For this purpose, the data must be transformed to the frequency domain on a period-by-period basis. The following estimates can then be made

$$\overline{V}(j\omega_k) = \frac{1}{M}\sum_{m=1}^{M} V_m(j\omega_k) \qquad (3-24)$$

$$\sigma_V^2(\omega_k) = \frac{1}{M-1}\sum_{m=1}^{M} \left(V_m(j\omega_k) - \overline{V}(j\omega_k) \right)\left(V_m(j\omega_k) - \overline{V}(j\omega_k) \right)^* \qquad (3-25)$$

where $V(j\omega_k)$ is interchangeably $U(j\omega_k)$ or $Y(j\omega_k)$. Since the Fourier coefficients are complex the variance calculated according to equation (3-25) will be the sum of the variances of the real and imaginary parts. The covariance can be estimated as

$$\sigma_{UY}(j\omega_k) = \frac{1}{M-1}\sum_{m=1}^{M} \left(U_m(j\omega_k) - \overline{U}(j\omega_k) \right)\left(Y_m(j\omega_k) - \overline{Y}(j\omega_k) \right) \qquad (3-26)$$

A significant covariance indicates that the input noise is not simply present in the input measurement channel but also passes through the system and is hence correlated with the output noise. This can be caused by noise on the measured input variable, such as turbulence in a measured flow, or by the presence of a feedback loop. The variances and covariance calculated in this way are double the quantities required for equations (2-34) and (2-37), where the variances and covariance of the real or imaginary parts are used.

It has been shown that the estimates of the sample variances and covariance are strongly consistent, for an increasing number of data points in each experiment, if the number of independent measurements M is greater than or equal to four (Schoukens *et al.*, 1997). In this work, each period of the test signal is considered as an independent measurement.

Since the estimation will be conducted only at the excited frequencies it is now possible to exclude all the other frequency points, which are termed *noise lines*, from the data.

Assuming the presence of stationary white noise, the improvement in SNR will be in the order of

$$\mathrm{SNR}_{ex} = \mathrm{SNR}_{td} + 10\log_{10}\left(\frac{N}{2F}\right) \tag{3-27}$$

which depends on the degree of over-sampling employed. Since the means and variances estimated from equations (3-24) and (3-25) are a measure of the periodic and stochastic power, respectively, they can be used to calculate the actual SNR following the exclusion of the noise lines

$$\mathrm{SNR}_{Vex} = 10\log_{10}\left(\frac{\sum_{k=1}^{F} |\overline{V}(j\omega_k)|^2}{\sum_{k=1}^{F} \sigma_V^2(\omega_k)}\right) \tag{3-28}$$

A further improvement in the SNR's is provided by averaging the data across M periods. The variance of the sample means is smaller than the sample variances estimated using equation (3-26), by a factor of $1/M$. This will provide a further improvement in the SNR of

$$\mathrm{SNR}_{Vav} = \mathrm{SNR}_{Vex} + 10\log_{10}(M) \tag{3-29}$$

3.4.4 Nonlinearities

The detection, analysis and modelling of nonlinear effects will be dealt with in detail in Chapters 4, 5 and 6.

3.5 Comparison of Signals

A comparison will now be made of the test times associated with single sine and multisine experiments, which will be followed by a discussion of the design criteria for MLBS signals.

3.5.1 Single Sines versus Multisines

Single sines are extensively used for frequency response measurements and the test times involved are often extended, due to the need to allow a settling time after the application of each sinusoid. This has fuelled an increased interest in the use of multisine signals, particularly in applications where the cost of testing is very high. Multisine tests can undoubtedly be made shorter but insufficient attention has been paid to the loss of accuracy which may be involved.

The total test time for single sines was given as equation (3-7) and for a multisine as equation (3-12). Their relative test time can be expressed as

$$\text{Relative } T_{\text{test}} = \frac{\text{Single Sine } T_{\text{test}}}{\text{Multisine } T_{\text{test}}} = \frac{(M\,T_0 + T_{\text{set}})F}{M\,T_0 + T_{\text{set}}} \qquad (3-30)$$

where T_0 is the period of the fundamental frequency. It is assumed that averaging is performed over the basic interval T_0 in each test. The single sine tests will take F times longer than those with the multisines and the relative benefit will increase with the number of test frequencies. However, this comparison takes no account of the relative accuracy of the tests. The bias and variance of FRF measurements were discussed in Chapter 2, where it was stated that the H_{EV} estimate is asymptotically unbiased with an uncertainty which can be approximated by equation (2-34). This means that the variance of the estimated transfer function is inversely proportional to $M\,|\,\overline{U}(j\omega_k)\,|^2$.

Hence, the relative accuracy will depend on the number of averages and the power of the input harmonics. The absolute accuracy will also depend on the noise power referred to the system output, which depends in turn on the nature of the noise and the system under test. For sampled data it will also depend on the cut-off frequency of the AA filters and the degree of over-sampling employed. To facilitate a comparison it will be assumed that the noise sequences disturbing the input and output are white and uncorrelated and that the total noise power is equal in each test.

Under these assumptions, the accuracy achieved with single sine tests will be the best possible, for a given input amplitude and number of averages. Consider the implications of achieving the same relative accuracy when using multisine signals. Assuming that the multisine has a flat amplitude spectrum, then the following equality must hold

$$\frac{M_{ms}}{M_{ss}} = \frac{|\,U_{ss}(j\omega)\,|^2}{|\,U_{ms}(j\omega)\,|^2} \qquad (3-31)$$

where M_{ms} is the number of averages of the multisine signal, M_{ss} the number of averages of the single sines, $|U_{ss}(j\omega)|$ the spectral amplitude of the single sines and $|U_{ms}(j\omega)|$ the spectral amplitude of the multisine at each test frequency.

It was observed in Section 3.3.4 that for $F > 30$, the CF of a consecutive or odd harmonic multisine is approximately equal to that of a single sine. Hence, for equal input amplitudes the total signal power will be approximately the same. This power will be concentrated at one frequency for the single sine but spread across F frequencies for the multisine. Thus, in order to maintain the same relative accuracy with multisines

$$M_{ms} \approx \frac{|\,U_{ss}(j\omega)\,|^2}{|\,U_{ss}(j\omega)\,|^2\,F^{-1}} M_{ss} \approx F\,M_{ss} \qquad \forall\ F > 30 \qquad (3-32)$$

which shows that the number of averages for a multisine will be greater by a factor of F. Substituting this relation back into equation (3-30) the relative test time can now be expressed as

$$\text{Relative } T_{\text{test}} = \frac{\text{Single Sine } T_{\text{test}}}{\text{Multisine } T_{\text{test}}} = \frac{M_{ss}\,T_0\,F + T_{\text{set}}\,F}{M_{ss}\,T_0\,F + T_{\text{set}}} \qquad (3-33)$$

which shows that the single sine tests will still take an additional time of $(F\text{-}1)T_{\text{set}}$ compared to using a multisine. The benefit of multisine testing is now seen to derive entirely from the reduction in settling time. This result can be verified by considering the design of signals to measure the dynamics of a gas turbine, which will be discussed in detail in Chapters 7 to 9.

Illustrative Example. The fuel feed to shaft speed dynamics of a gas turbine are a practical example of an overdamped multiple-pole system. The low pressure shaft dynamics have s-domain poles at s_1 = -0.45 and s_2 = -1.8, with corresponding time constants $\tau_1 = 2.2$ s and $\tau_2 = 0.55$ s. In order to measure the FRF of this system a frequency range must be selected which adequately covers the system break-points. Since the break-points are at 0.07 Hz and 0.29 Hz then a frequency range of 0.01-0.6 Hz is seen to be adequate.

A 30 odd harmonic multisine with a fundamental frequency of 0.01 Hz has a bandwidth of 0.01 - 0.59 Hz and the $L\infty$ compression of this signal results in a CF of 1.35 and a TF of 0.92. The signal is shown in the time and frequency domains in Figure 3-5. Selecting a settling time of eight times the slowest time constant, the minimum test times for single sine and multisine signals will be

$$\text{Single Sine} \quad T_{\text{test}} = \left(\frac{1}{0.01} + 8 \times 2.2 \right) \times 30 = 3528\, s \tag{3-34}$$

$$\text{Multisine} \quad T_{\text{test}} = \frac{1}{0.01} + 8 \times 2.2 = 117.6\, s \tag{3-35}$$

which shows that the multisine tests can be made considerably shorter.

Considering the relative accuracy, it can be seen from Figure 3-5 that for a maximum time-domain amplitude of ±10 units the multisine harmonics each have an amplitude of 1.9 units. The equivalent single sines will each have an amplitude of 10 units. Hence from equation (3-31)

$$M_{ms} = \frac{10^2}{1.9^2} \approx 28\, M_{ss} \tag{3-36}$$

The multisine test time required to achieve the same accuracy as the single sines will be

$$\text{Multisine} \quad T_{\text{test}} = 28 \times \frac{1}{0.01} + 8 \times 2.2 = 2818\, s \tag{3-37}$$

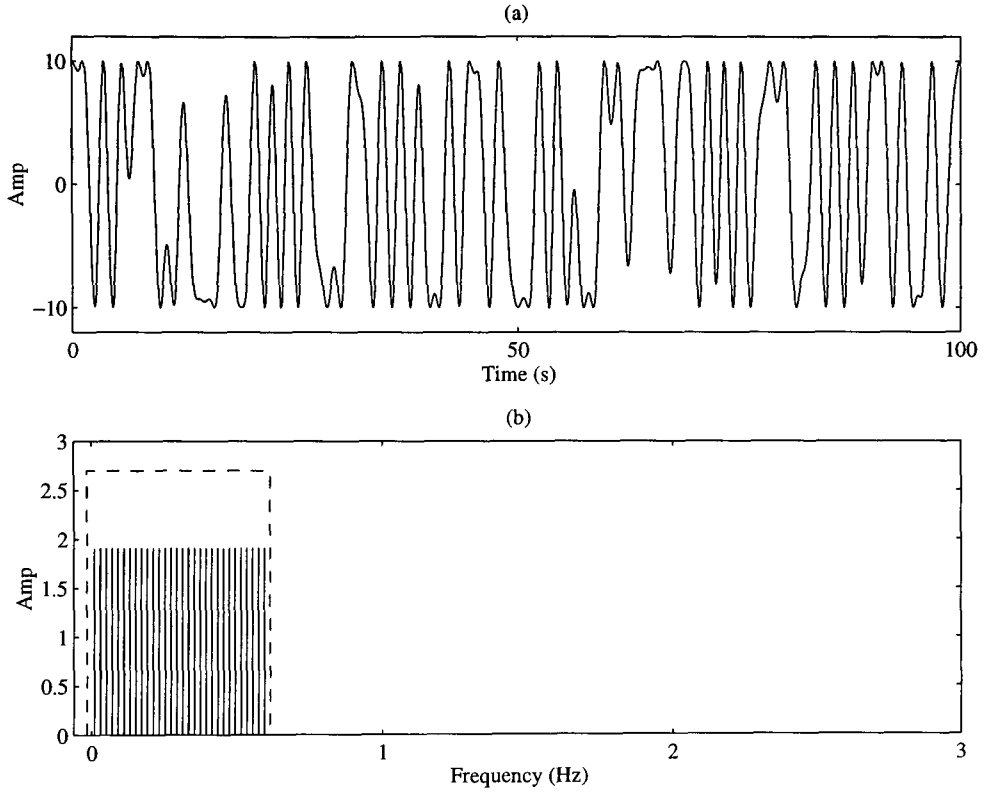

Figure 3-5. Multisine signal in (a) time domain and (b) frequency domain, with the frequency range of the system poles shown boxed.

Comparison with equation (3-34) shows that a 20% reduction in test time can still be achieved. This reduction is mainly due to the reduction in settling time but is also due to the multisine CF being slightly lower than that of the single sines, which emphasises the importance of CF minimisation.

In practice, if the input and output SNR's are high, then it may not be necessary to carry out so many averages of the multisine in order to achieve an acceptable level of accuracy. A further reduction in test times can then be achieved. In the case of the gas turbine data, discussed in detail in Chapter 8, an acceptable accuracy was achieved after only six averages. This allowed the test times to be reduced by around 80%. The reduction of 20% can thus be thought of as the minimum achievable in this example by using multisine tests.

Finally, it is interesting to examine the use of the TF factor as a discriminatory measure between single sine and multisine signals. The TF, as defined in equation (3-5), is

$$\mathrm{TF} = \max\left\{\frac{u_{\max}^2}{F\,|\,U(j\omega_k)\,|^2}\right\} \qquad k = 1, 2, \ldots, F \tag{3-38}$$

and for both single sines and flat-amplitude multisines the denominator term is simply the total power in the signal. Using Parseval's equation the TF can thus be expressed as

$$\mathrm{TF} = \frac{u_{\max}^2}{2u_{rms}^2} = \frac{\mathrm{CF}^2}{2} \tag{3-39}$$

which is simply half the CF squared and provides no additional information about the relative test times! This shows that while the TF is an useful tool for comparing multifrequency signals with non-flat spectra, where its frequency selectivity allows the presence of weak frequency components to be penalised, it is not an useful measure in this case. Indeed, this result simply emphasises once again the importance of minimising the CF of the multisine signals.

3.5.2 MLBS Design

The time-domain design criteria for MLBS signals were given as equations (3-16) and (3-17) in Section 3.3.5. The criteria must be extended for systems with more than one time constant and for the purpose of gas turbine testing Hill (1994), following Moore (1970) and Cottington (1979), specified that

$$\Delta t = \frac{\tau_{\min}}{10} \qquad \text{and} \qquad N\,\Delta t = 10\,\tau_{\max} \tag{3-40}$$

which combine to give

$$N = 100\,\frac{\tau_{\max}}{\tau_{\min}} \tag{3-41}$$

An alternative frequency-domain approach is proposed, whereby the half power point of the MLBS spectrum is placed close to the maximum frequency of interest by adjusting Δt.

$$\Delta t = \frac{0.443}{f_{\max}} \qquad (3-42)$$

The sequence length N can then be varied to set the number of harmonics within the frequency range of interest, as shown in Table 3-2. It can be seen that the number of harmonics up to the half power point is approximately half the sequence length. Table 3-2 also shows that the greater the number of harmonics, the lower the amplitude of the fundamental and each subsequent harmonic.

Illustrative Example. The differing designs can be illustrated using the gas turbine model discussed in the previous section. The time-domain criteria specify an MLBS sequence with

$$N = 100\frac{2.2}{0.55} = 400 \qquad \Delta t = \frac{0.55}{10} = 0.055 \qquad (3-43)$$

to which the nearest appropriate sequence length is $N = 512$. The spectrum of an inverse repeat sequence based on this design is shown in Figure 3-6 (a), for a time-domain amplitude of ± 10 units of fuel flow, along with a boxed area denoting the frequency range of the system poles.

It is evident from the plot that the signal has far too high a bandwidth, with the vast majority of the signal power concentrated outside the frequency range of interest. It is also clear that the signal power is spread among too many harmonics, which results in the low values of power per harmonic. If the first 30 harmonics were used to estimate the FRF then the amplitude of the 30th component would be only 0.9 units, which is less than half that of the multisine shown in Figure 3-5. This results in a TF of 4.3, which strongly indicates that the signal is not suitable for this purpose.

Applying the frequency-domain approach, if Δt is set to 0.79 s and N to 63, the resulting inverse repeat sequence will have 30 odd harmonics within the frequency range of interest. The spectrum of the signal is shown in Figure 3-6 (b) and it is clearly more

appropriate for the system dynamics than that of Figure 3-6 (a). The TF of this signal is only 1.1, indicating that it is far more suitable than the previous design. The TF is seen to be an useful measure for discriminating between different MLBS signals.

The spectrum of the second binary sequence can be compared with that of the multisine shown in Figure 3-5. The power in the 30th harmonic is 1.72 units, which is only slightly less than that of the multisine signal. Thus to achieve the same minimum accuracy of FRF estimation, the test times with the binary signal and odd harmonic multisine would be of comparable length. This is also shown by their relative TF, since the multisine has a TF of 0.9 compared with 1.1 for the MLBS signal.

The much reduced power at certain points in the spectra of the inverse repeat sequences is a feature of the design. It has been shown that this effect occurs at harmonics which are odd multiples of the original MLBS sequence length (Godfrey, 1993b, pp. 92-95). The signal shown in Figure 3-6 (b) was based on an MLBS sequence length of 63 and hence the dips in amplitude occur at the 63rd and 189th harmonics.

TABLE 3-2

EFFECT OF VARYING MLBS SEQUENCE LENGTH

Shift Register Stages (n)	Sequence Length (N)	Harmonics up to $\approx$ -3 dB Point	Amplitude of Fundamental (with $V = \pm 1$)
4	15	8	0.26
5	31	15	0.18
6	63	30	0.13
7	127	60	0.09
8	255	120	0.06
9	511	240	0.04
10	1023	480	0.03
11	2047	960	0.02

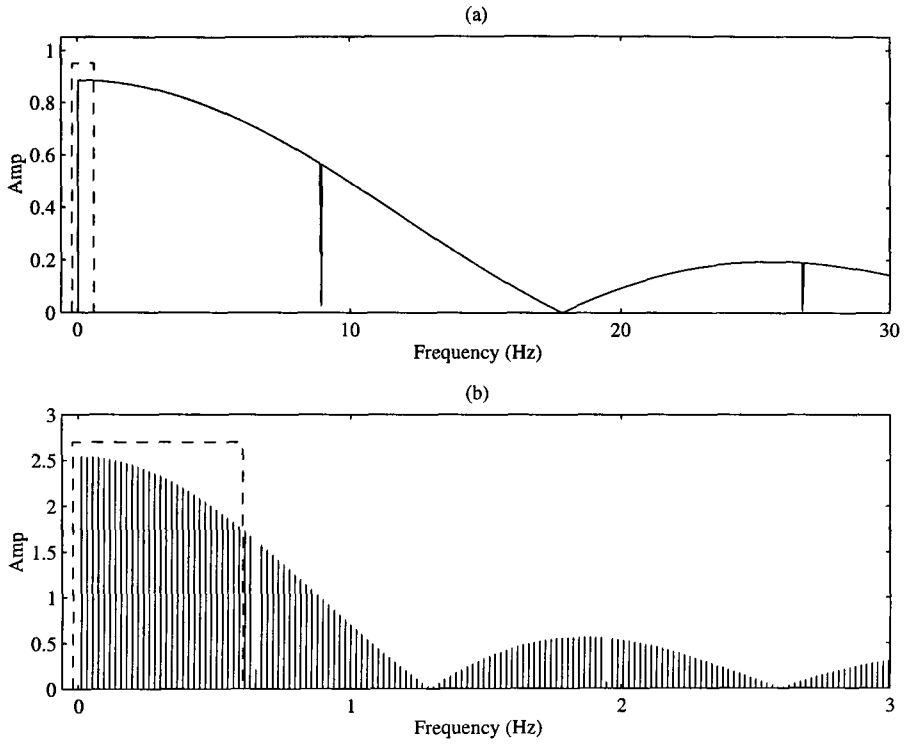

Figure 3-6. Inverse repeat MLBS designs, based on (a) time-domain criteria and (b) frequency-domain criteria, with the frequency range of the system poles shown boxed.

3.6 Conclusions

An overview has been presented of the main issues relating to the design of the identification experiment, with particular emphasis on the choice of input signals. The implications for the experimental set-up of the basic assumptions underlying continuous and discrete modelling have been emphasised. The advantages of periodic signals for both continuous and discrete modelling have been presented and the properties of single sines, multisines and MLBS signals have been discussed.

Two measures to assess signal quality have been dealt with, the crest factor and time factor. The CF simply measures the amplitude compression of a signal in the time domain, relative to its RMS value. The TF aims to provide a measure of the test time required to obtain a minimum accuracy of FRF estimation and hence penalises any weak components in the signal spectrum.

The CF minimisation of multisine signals was discussed in detail and the relative merits of the different approaches were outlined. Of the methods discussed, the lowest CF are obtained with the $L\infty$ method proposed by Guillaume (1991). It was shown that the CF's of flat-amplitude multisines of 30 or more consecutive or odd harmonics can be made equal to, or slightly less than, the CF of a single sine. Thus the single sine and multisine signals will have the same total power, for a given input amplitude constraint.

The relative test times for FRF measurements using single sine and multisine signals was then examined in some detail. It was shown that the test times can be made considerably shorter using multisine signals if the effect on the accuracy is ignored. In order to address this, the relative test times were compared under the constraint that the relative accuracy of the single sine and multisine tests should be equal. It was shown that the number of averages of a multisine signal required to equal the accuracy of single sine tests is approximately the number of frequencies in the signal. The reduction in test time associated with multisines in then derived entirely from the reduction in settling time.

These results were illustrated using the practical example of test signal design for gas turbine modelling. The reduction in test time achieved using the multisine signal was shown to derive mainly from the reduction in settling time. A further reduction was also achieved in this example because the CF of the multisine was slightly less than that of the single sines, which emphasises the importance of CF minimisation.

It was pointed out that if the SNR's are good then less averages of the multisine may be necessary in order to achieve an acceptable accuracy. The benefit derived from the reduction in settling times should therefore be considered as the minimum achievable using multisine signals. It was also shown that for single sines and flat-amplitude multisines the TF reduces to a value of half the squared CF and hence provides no more information than the CF itself. This result further emphasises the importance of minimising the multisine CF.

The design of MLBS signals was discussed and the problems with the time-domain design criteria commonly employed were illustrated. An alternative frequency-domain approach was proposed, based on placing the half power point of the MLBS spectrum close to the maximum frequency of interest and then adjusting the sequence length to place an appropriate number of frequencies within that range. An MLBS signal designed in this way was shown to have comparable spectral properties to an $L\infty$ multisine, in the

frequency range of interest, and hence comparable test times. It was also shown that the TF is useful in discriminating between different MLBS designs, since it acts to penalise any weak frequency components.

Multisine signals are preferred over binary signals due to their flexibility, in that they allow a completely arbitrary input spectrum to be defined. Multisines will therefore be the signals considered for the remainder of this thesis.

Chapter IV

Nonlinear Systems

Abstract — *This chapter examines the effects of power series nonlinearities on periodic multifrequency signals. A novel methodology is proposed which provides new insights into the differing effects of even- and odd-order nonlinear terms. It is shown that odd-order nonlinearities will always generate contributions at the test frequencies, whatever the harmonic content of the signal. A class of broad-band pilot test signals is proposed, termed sparse odd multisines, which can be used to establish the system bandwidth and detect nonlinearities. Signals are then defined within this class which allow the measurement of the best linear approximation to a nonlinear system.*

4.1 Introduction

Linear system identification involves approximating complex reality with relatively simple models. While the exact equivalence of model and reality is impossible, the model should match the real system as closely as possible within the constraints of test time, computation time and model complexity. Much attention has been paid to the influence of stochastic effects on the estimation of linear models, far less has been paid to the effects of system nonlinearities. Yet all practical systems are nonlinear to some extent.

The nonlinear systems considered in this thesis will be restricted to those which can be modelled by series expansions. This chapter begins with a review of *block-oriented* models, which are composed of dynamic linear and static nonlinear elements. A novel analysis method is presented for studying the contributions generated by static polynomial nonlinearities subjected to multifrequency inputs, which is then extended to nonlinear systems with memory. It is shown that odd-order nonlinearities will always generate contributions at the input frequencies, whatever the frequency content of the test signal.

The detection of nonlinear effects using broad-band signals is then addressed and a class of signals, termed sparse odd multisines, is defined. A new type of multisine is then designed within this class, which suffers from only one type of nonlinear distortion, for a given order of nonlinearity. Application of such a signal reduces the nonlinear influence to that of a gain term and it can thus be used to measure the best linear approximation to the overall nonlinear system. This concept is discussed in relation to the work of other authors in this field, who use a different approach to make the same measurement.

The analysis presented in this chapter also contributes to a better understanding of the means by which the nonlinear distortion can be minimised, which will be discussed in Chapter 5 and the design of signals to measure Volterra kernels, presented in Chapter 6.

4.2 Nonlinear Systems

Consider a periodic multisine signal applied to a time invariant system. Any nonlinearities present will generate an output contribution which will be the same for each

successive period of the signal. This will introduce a systematic distortion into the estimated linear FRF which, in contrast to the error introduced by stochastic effects, will not reduce with averaging

$$\hat{H}(j\omega_k) = \frac{\overline{Y}(j\omega_k)}{\overline{U}(j\omega_k)} = \frac{\overline{Y}_l(j\omega_k) + \overline{Y}_{nl}(j\omega_k)}{\overline{U}(j\omega_k)} \qquad (4-1)$$

where $Y_l(j\omega_k)$ is the linear response and $Y_{nl}(j\omega_k)$ the nonlinear distortion at the test frequencies. It will be shown that this introduces a bias into the estimated linear FRF and the parametric s-domain models. The severity of the distortion can be quantified by the mean magnitude of the complex error between the linear FRF and the estimated FRF. This quantity is termed the Chebyshev error and is used as a basis for filter design, since it takes account of distortion in both amplitude and phase (Chen and Parks, 1987). It will be referred to simply as the *complex error* for the remainder of this work.

$$E = \frac{1}{F} \sum_{k=1}^{F} | H(j\omega_k) - \hat{H}(j\omega_k) | \qquad (4-2)$$

The nature of this distortion will now be examined for a range of block-oriented models, composed of linear systems in various combinations with a static nonlinearity, where the nonlinear element is defined by

$$N(x(t)) = \sum_{n=1}^{\infty} a_n x(t)^n \qquad (4-3)$$

The parallel nonlinear structure shown in Figure 4-1 (a) is the most basic nonlinear model, composed of a linear system in parallel with a static nonlinear element (Bendat, 1990, chapter 2). The Hammerstein model of Figure 4-1 (b) was first postulated by Narendra and Gallman (1966) and has been extensively studied since then (Greblicki and Pawlak (1986)). The Wiener model shown in Figure 4-1 (c) is a reduced form of the general Wiener model, described by Schetzen (1980), which has been studied in detail by Korenberg (1991).

Finally, the Wiener-Hammerstein model of Figure 4-1 (d), also referred to as the cascade model, has been studied by a large number of authors. The analysis and synthesis of this structure was addressed by Shanmugam and Lal (1976) and by Palm (1978). A useful

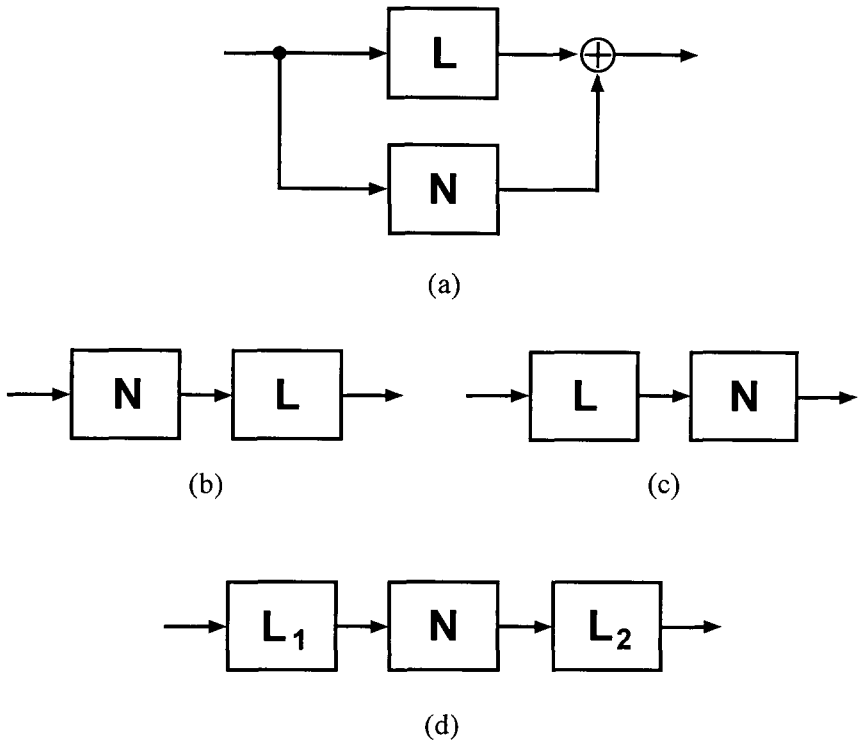

Figure 4-1. Nonlinear model structures: (a) Parallel, (b) Hammerstein, (c) Wiener and (d) Wiener-Hammerstein or Cascade.

overview of the work conducted on the structure was presented by Billings and Fakhouri (1979), while Billings and Tsang (1990) presented a frequency-domain analysis of the model, based on correlation methods using white noise as an input. The uniqueness of the model, in terms of input and output measurements, has been established by Boyd and Chua (1983). The structure is an attractive compromise between the simplicity of the other block-oriented models and the complexity of alternative approaches to nonlinear modelling.

While the overall nonlinear effect will vary for each model, it will depend in each case on the signal applied to the static nonlinear element. The analysis which follows is thus applicable to each of these model structures. The effect of even- and odd-order static nonlinearities on multisine inputs will first be examined, with the analysis being extended in a straightforward manner to nonlinear systems with memory in Section 4.2.3.

The distortions introduced by a power series nonlinearity are the product of time-domain multiplication and hence convolution in the frequency domain. Defining the discrete frequency-domain convolution as

$$U(j\omega)*U(j\omega) = \sum_{\lambda} U(j\lambda)\, U(j(\omega-\lambda)) \qquad (4-4)$$

The output from a static cubic nonlinearity can be expressed as

$$Y_3(j\omega) = a_3\, [U(j\omega)*U(j\omega)*U(j\omega)] \qquad (4-5)$$

where a_3 is the coefficient of the cubic term. If the input is a single cosine of amplitude $2A$, with a double sided spectrum

$$U(j\omega) = A[\, \delta(\omega+\omega_0) + \delta(\omega-\omega_0)\,] \qquad (4-6)$$

the output from the cubic nonlinearity will be

$$\begin{aligned} Y_3(j\omega) &= a_3 A^3 [\, \delta(\omega-(\omega_0+\omega_0+\omega_0)) + \delta(\omega-(-\omega_0-\omega_0-\omega_0)) \\ &\quad + \delta(\omega-(-\omega_0-\omega_0+\omega_0)) + \delta(\omega-(-\omega_0+\omega_0-\omega_0)) + \delta(\omega-(\omega_0-\omega_0-\omega_0)) \\ &\quad + \delta(\omega-(\omega_0+\omega_0-\omega_0)) + \delta(\omega-(\omega_0-\omega_0+\omega_0)) + \delta(\omega-(-\omega_0+\omega_0+\omega_0))] \\ &= a_3 A^3 [\, \delta(\omega+3\omega_0) + 3\delta(\omega+\omega_0) + 3\delta(\omega-\omega_0) + \delta(\omega-3\omega_0)\,] \end{aligned} \qquad (4-7)$$

consisting of three contributions at ω_0 and one at $3\omega_0$, which result from the various permutations, with repetition, of combinations of three of the input harmonics. If the signal is a multisine of F cosines, with dc excluded

$$U(j\omega) = \sum_{\substack{k=-F \\ k\neq 0}}^{F} a(k)\, e^{j\phi(k)}\, \delta(\omega - i(k)\omega_0) \qquad (4-8)$$

and

$$\begin{aligned} a(-k) &= a(k) \\ i(-k) &= -i(k) \qquad\qquad k = 1 \ldots F \\ \phi(-k) &= -\phi(k) \end{aligned}$$

then the output from the cubic element will be

$$\begin{aligned} Y_3(j\omega) = a_3 \sum_{\substack{p=-F \\ p\neq 0}}^{F} \sum_{\substack{m=-F \\ m\neq 0}}^{F} \sum_{\substack{k=-F \\ k\neq 0}}^{F} & [a(p)\, a(m)\, a(k)] \\ & \times \exp\left(\, j(\, \phi(p)+\phi(m)+\phi(k)\,)\,\right)\ \delta\left(\omega - [i(p)+i(m)+i(k)]\omega_0\right) \end{aligned} \qquad (4-9)$$

and the output will contain $8F^3$ nonlinear contributions. By extension, the output from a nonlinear term of order n will consist of $(2F)^n$ nonlinear contributions, made up of all possible combinations of n input harmonics.

$$Y_n(j\omega) = a_n \sum_{\substack{m_1=-F \\ m_1\neq 0}}^{F} \sum_{\substack{m_2=-F \\ m_2\neq 0}}^{F} \cdots \sum_{\substack{m_n=-F \\ m_n\neq 0}}^{F} \left(\prod_{p=1}^{n} \mathbf{a}(m_p) \right) \exp\left(j \sum_{p=1}^{n} \phi(m_p) \right) \delta\left(\omega - \left(\sum_{p=1}^{n} i(m_p) \right) \omega_0 \right) \qquad (4-10)$$

It is common practice to consider the contributions generated by an nth-order nonlinearity as two types: the *nth harmonics*, generated by sums of the same input harmonics, and *cross-talk* or *intermodulation* harmonics, made up of sums of different input harmonics. This distinction owes its origin to communications engineering (Stremler, 1982, chapter 7) and clearly has some value when considering inputs made up of only a few harmonics. However, it can lead to misunderstanding when studying the contributions generated by multifrequency signals, since the nth harmonics are only a very small number of the total nonlinear contributions. The following methodology is proposed in order to clarify the influence of power series nonlinearities on such signals.

4.2.1 Even-Order Terms

It is clear from the summation of impulses in equation (4-10), that if n is even and the input signal only contains harmonics which are odd multiples of the fundamental, then all the output components will fall at even harmonics. Thus, by using an odd harmonic multisine

$$u(t) = \sum_{k=1}^{F} a(k)\cos(i(k)\omega_0 t + \phi(k)) \qquad \mathbf{i} = [\,1 \quad 3 \quad 5 \quad 7 \quad \ldots \quad (2F-1)\,] \qquad (4-11)$$

the linear and nonlinear contributions at the output will be orthogonal in the frequency domain. The even harmonics can then be omitted from the data set used for estimation. Such signals are the multisine equivalent of inverse repeat binary sequences, which have been employed in cross-correlation analysis to reduce the influence of predominantly second-order nonlinearities (Godfrey and Moore (1974) and Godfrey (1993a)).

4.2.2 Odd-Order Terms

If n is odd, and only odd harmonics are included in the input signal, then only odd harmonics will be generated by the nonlinear term. Thus the components generated by

odd- and even-order terms will be orthogonal in the frequency domain. The odd-order nonlinear terms will also generate contributions at the odd excitation harmonics, which can be divided into two categories.

Type I Contributions. These are contributions which fall only at the test frequencies and are present whatever the frequency content of the signal. They are generated by combinations of a test frequency with pairs of equal positive and negative frequencies, which results in a contribution at the test frequency itself.

For example, for a cubic nonlinearity, the combinations $\delta(\omega-(\omega_0-\omega_0+\omega_0))$ and $\delta(\omega-(\omega_0+3\omega_0-3\omega_0))$ will each produce output contributions at ω_0. The first example shows the combination of an input frequency with its own negative and positive frequency pair, while the second example shows the combination of an input frequency with a positive and negative frequency pair of another frequency.

Equation (4-12) shows the contributions generated at a given excitation frequency ω_k, for any odd-order n, where C_p is the number of possible combinations of $\mathbf{i}(k)\omega_0$ with $(n\text{-}1)/2$ complex conjugate frequency pairs. It can be seen that two terms within the expression go to zero, indicating that the complex conjugate pairs in the combination have no effect on the phase or frequency of the resulting contribution.

$$T_1(j\omega_k)=a(k)\sum_{m_1=1}^{F}\sum_{m_2=1}^{F}\cdots\sum_{m_R=1}^{F}C_n(m_1)\prod_{p=1}^{R}a^2(m_p)$$
$$\times\exp\left(j\phi(k)+j\sum_{m_1=1}^{F}\cdots\sum_{m_R=1}^{F}\sum_{p=1}^{R}(\phi(m_p)-\phi(m_p))\right)$$
$$\times\delta\left(\omega-\left(i(k)\omega_0+\sum_{m_1=1}^{F}\cdots\sum_{m_R=1}^{F}\sum_{p=1}^{R}(i(m_p)\omega_0-i(m_p)\omega_0)\right)\right)\qquad R=\frac{n-1}{2}$$
(4-12)

These Type I contributions will always have the same phase as the original test frequency. The number of contributions generated is equal for each test frequency and depends only on the order of the nonlinearity and the number of frequencies in the input multisine. Altering the specific frequencies which are included in the signal will in no way affect the number of these *harmonic* contributions. They will therefore tend to introduce a bias, in the form of a systematic *offset* at each frequency of the measured output signal, which is proportional to the input amplitude at that frequency.

Type II Contributions. These contributions fall at both the test frequencies and frequencies omitted from the input signal. They depend entirely on the harmonic vector **i** and the number falling at the test frequencies will vary greatly from one frequency to another. For example, for a cubic nonlinearity, the combinations $\delta(\omega-(5\omega_0-9\omega_0+7\omega_0))$ and $\delta(\omega-(7\omega_0-3\omega_0-1\omega_0))$ will each produce contributions at $3\omega_0$. The contributions generated at a given frequency ω_k by any odd-order n are shown in equation (4-13). Frequency combinations which can be grouped as ω_k plus pairs of equal positive and negative frequencies are explicitly excluded from this set, since they correspond to the Type I contributions described in equation (4-12).

$$T_2(j\omega_k)=\sum_{\substack{m_1=-F\\ m_1\neq 0}}^{F}\sum_{\substack{m_2=-F\\ m_2\neq 0}}^{F}\cdots\sum_{\substack{m_n=-F\\ m_n\neq 0}}^{F}\left(\prod_{p=1}^{n}a(m_p)\right)$$

$$\times\exp\left(j\sum_{p=1}^{n}\phi(m_p)\right)\delta\left(\omega-i(k)\omega_0\,\varepsilon\left(i(k)\omega_0-\sum_{p=1}^{n}i(m_p)\omega_0\right)\right) \qquad (4-13)$$

where $\varepsilon(x)=\begin{cases}1 \text{ if } x=0\\ 0 \text{ if } x\neq 0\end{cases}$

The phase of these contributions will vary depending on the phases of the specific combination of frequencies which gave rise to them. Omitting certain harmonics from the signal will restrict the possible number of these *inter-harmonic* contributions which will fall at a given frequency. Since the number and relative phase of these contributions will vary from frequency to frequency the resultant bias will also vary. These contributions will introduce a bias in the form of a *scatter*, which might easily be mistaken for a stochastic effect.

On the basis of these definitions, the harmonics generated by even-order nonlinear terms can also be classified as Type I and Type II contributions. The Type I contributions will all fall at dc in this case, since they will be generated by pairs of equal positive and negative harmonics. All other contributions generated will be Type II, since altering the harmonic content of the signal will directly influence their values.

This analysis method presents a clearer picture of the mechanisms at work in the generation of nonlinear contributions than their conventional classification into *n*th harmonics and intermodulation harmonics. It clearly shows which type of odd-order contributions can be influenced by the selection of the input harmonics and which type cannot and also points to the differing influence of the two types on the output spectrum.

4.2.3 Nonlinear Systems with Memory

The analysis can be extended to nonlinear systems with memory by including an additional term in equation (4-10).

$$Y_n(j\omega) = a_n \sum_{\substack{m_1=-F \\ m_1\neq 0}}^{F} \sum_{\substack{m_2=-F \\ m_2\neq 0}}^{F} \cdots \sum_{\substack{m_n=-F \\ m_n\neq 0}}^{F} H_n(j\omega_0[i(m_1), i(m_2), \ldots, i(m_n)])$$

$$\times \left(\prod_{p=1}^{n} a(m_p) \right) \exp\left(j \sum_{p=1}^{n} \phi(m_p) \right) \delta\left(\omega - \left(\sum_{p=1}^{n} i(m_p) \right) \omega_0 \right) \qquad (4-14)$$

The term $H_n(\bullet)$ is a complex weighting of n dimensions, which is applied to each of the combinations generated by the nonlinearity. This is a Volterra type expression, where the system output is composed of the summation of outputs from a first-order linear kernel and higher order kernels, which will be dealt with in greater detail in Chapter 6. The number and type of contributions generated at each frequency will be the same as for the static case, except that their magnitude and phase will be influenced by $H_n(\bullet)$.

4.3 Detecting the Nonlinearity

Before proceeding with testing and identification it is necessary to establish whether the system can be considered linear across the specified time-domain input amplitude. There are a variety of techniques for establishing the linearity of a system, the simplest being to inject a series of single sinewaves of increasing amplitude and look for distortion of the output signal (Haber, 1985).

Some idea of the system bandwidth is also required in order to properly design the test signals for linear identification. This can be achieved by injecting a wide-band pilot test signal. However, time is usually at a premium when testing industrial systems, due to operational or financial constraints. It is therefore desirable to minimise the time spent on initial tests by combining these two operations. This can be achieved by using a wide-band multisine, with certain harmonics excluded, as the pilot signal.

If all the even harmonics and some of the odd harmonics are excluded from the signal then nonlinear effects can be detected by the presence of Type II harmonics between the test frequencies in the output. Additional Type II harmonics will also fall at frequencies

higher than the signal bandwidth, but these may be attenuated if the unknown system under test has a much smaller bandwidth than the wide-band pilot signal. The nonlinear effect might then be missed if a consecutive harmonic signal was used.

A criterion is required to assess whether any of the additional output components are indeed periodic Type II contributions, or simply noise harmonics. McCormack *et al.* (1995) proposed using the squared coherence function, defined for systems with a noise free input, which depends only on the output signal

$$\gamma_{nl}^2(\omega) = \frac{|\frac{1}{M}\sum_{m=1}^{M} Y_m(j\omega)|^2}{\frac{1}{M}\sum_{m=1}^{M} Y_m(j\omega)Y_m^*(j\omega)} = \frac{|\bar{Y}(j\omega)|^2}{G_{YY}(\omega)} \qquad (4-15)$$

where $Y_m(j\omega)$ is the output spectrum at the excited and nonexcited frequencies, measured across M periods. The quantity $\gamma_{nl}^2(\omega)$ will be termed the *nonlinear coherence* to distinguish it from the more commonly used definition of coherence, given in equation (2-29). Assuming the presence of uncorrelated input and output noise then equation (4-15) can be expressed as

$$\gamma_{nl}^2(\omega) = \frac{G_{YY}(\omega) + (G_{NN}(\omega)/M)}{G_{YY}(\omega) + G_{NN}(\omega)} \qquad (4-16)$$

which shows that, as M becomes large, the nonlinear coherence will express the ratio of the periodic power over the total power at each output frequency. If the periodic component is zero, it will assume a value of $1/M$, which gives an useful bound with which to judge the significance of the nonlinear coherence values. It can be assumed that all those values which lie close to this bound are simply a function of noise and do not indicate the presence of any periodic nonlinear contributions.

4.3.1 Sparse Odd Multisines

The design of multifrequency signals with spaces in their spectra has been studied by a number of authors. Barker (1993) proposed a method for designing *multilevel* (m-level) signals with certain harmonics and their multiples suppressed, which were applied by Barker and Al-Hilal (1985) to the identification of a Hammerstein model. An alternative approach to designing m-level signals with harmonic suppression was proposed by

McCormack *et al.* (1995). These m-level signals have a number of attractive properties, in particular the amount of power that can be injected at certain frequencies for a given time-domain amplitude. However, only multisines afford the user complete control over the input spectrum and they are preferred for this reason.

Multisines in which all even harmonics are excluded, along with some odd harmonics, will be termed *sparse odd* signals. Any number of odd harmonics may be excluded from a multisine signal in order to create spaces for nonlinear detection. The family of sparse odd multisines covers a range of signals, beginning with an odd multisine from which only a single odd harmonic has been excluded and arriving at a single sine, from which all other harmonics are excluded. If the aim is to detect the presence of a nonlinearity then the amplitude of the nonlinear contribution generated at the excluded harmonics by the nonlinearity should be maximised.

One option is to increase the input signal amplitude, since the aim is now to detect the nonlinearity, which implies that it must be sufficiently excited. But this may not be possible if the linear model being identified is only valid over a restricted input range. In the discussion which follows it is assumed that such an upper limit exists on the time-domain amplitude of the input signal.

The other variables affecting the response to a given sparse signal are the phases of the input harmonics and the number of excluded harmonics. While it is impossible to present general conclusions for all nonlinear systems, the following analysis will examine the effect of varying the signal phase and sparsity for a static cubic nonlinearity. To facilitate comparison, a *sparsity index* (SI) will be defined, which is zero for an odd harmonic signal and tends to unity as the sparsity is increased.

$$\text{Sparsity Index} = 1 - \frac{\text{Number of Included Harmonics}}{\text{Total Possible Odd Harmonics}} \qquad (4-17)$$

Five signals of increasing sparsity will be studied. Firstly, a signal will be designed in which the second and third harmonics and their multiples are excluded, which is the multisine equivalent of the m-level signals designed by Barker and Al-Hilal (1985) and McCormack *et al.* (1995). Suppressing the harmonics which are multiples of three means that one in three odd harmonics is excluded.

$$\text{Signal 1}: \quad \mathbf{i} = \left[1 \quad 5 \quad 7 \quad 11 \quad 13 \quad 17 \quad \ldots \quad \begin{cases} (3F-2) & F \text{ odd} \\ (3F-1) & F \text{ even} \end{cases} \right] \tag{4-18}$$

Secondly, a signal where every other odd harmonic is excluded, termed an *odd-odd multisine*, will be designed.

$$\text{Signal 2}: \quad \mathbf{i} = [1 \quad 5 \quad 9 \quad 13 \quad 17 \quad 21 \quad \ldots \quad (4F-3)] \tag{4-19}$$

Then a series of signals of increasing sparsity will be studied

$$\text{Signals 3/4/5}: \quad \mathbf{i} = [1 \quad \ldots \quad (2r(F-1)+1)] \qquad r = 3, 4, 5 \tag{4-20}$$

The total number of contributions generated by each of these signals at the excluded odd harmonics within the signal bandwidth is shown in Table 4-1, where signals of 15 input harmonics are compared. The total number of contributions is seen to increase slightly between signals one and two but then to remain constant, despite the increasing sparsity. This shows that the total number of contributions remains the same for the signals with a regular harmonic spacing.

The amplitude of the resultant nonlinear harmonics will depend on the interaction of these contributions, which in turn will depend on the phase of the input harmonics. Assuming that an upper limit exists on the time-domain excursion of the input signal, it was found that the maximum power at the excluded frequencies was generated with low CF signals.

TABLE 4-1

CONTRIBUTIONS GENERATED AT THE EXCLUDED ODD HARMONICS

Signal	Sparsity Index	Number of Contributions
1	0.32	2135
2	0.48	2240
3	0.65	2240
4	0.74	2240
5	0.79	2240

Illustrative Example. The results of applying an odd-odd multisine signal will now be illustrated by means of a simulation. The parallel nonlinear structure shown in Figure 4-1 (a) was employed, with a first-order linear system in parallel with a cubic nonlinear term. Gaussian noise was added to the system input and output and six periods of a 20 harmonic odd-odd signal were measured.

The time and frequency plots of the input and output of the system are shown in Figure 4-2. The time plots show the signals after averaging, while the frequency plots show the FFT calculated across the complete record lengths. No obvious nonlinear distortion can be observed from the time records but the frequency-domain picture is far more revealing. The input signal contains the excited harmonics, at close to 0 dB, and additional harmonics at -60 dB or less, which can be attributed to the noise. At the output, additional power is present at the omitted odd harmonics, which have risen above the noise floor to around -40 dB.

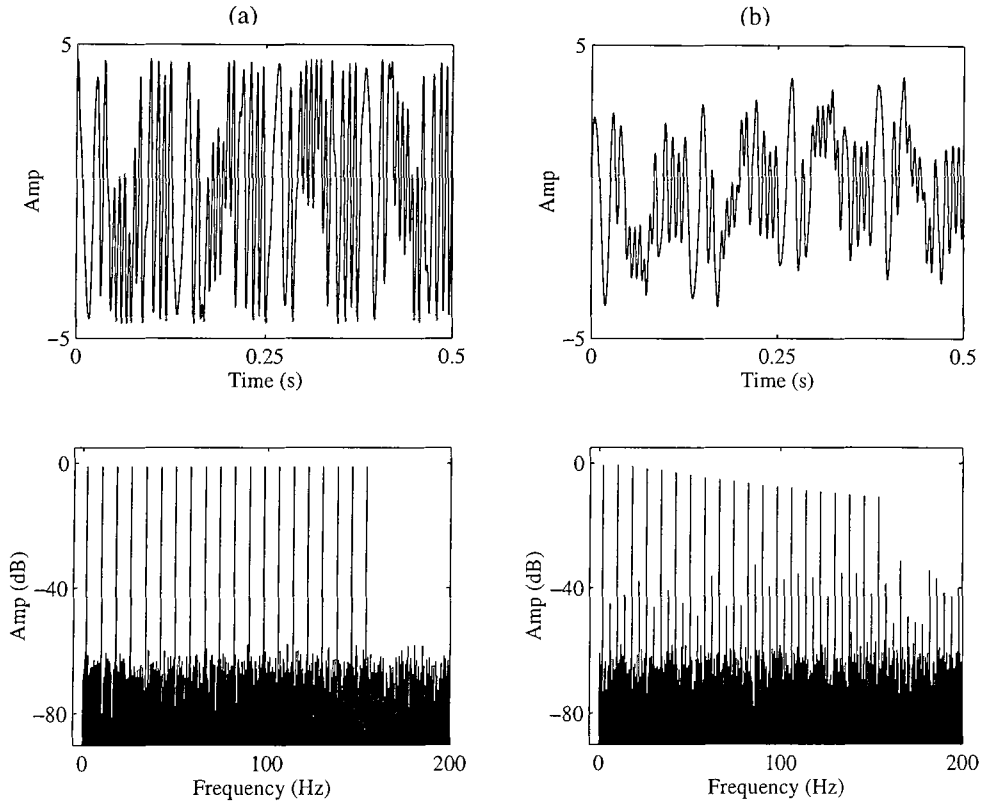

Figure 4-2. Plot of (a) odd-odd input signal and (b) system response.

A plot of the nonlinear coherence of the non-excited even and odd harmonics of the odd-odd signal is shown in Figure 4-3. The nonlinear coherence of the input signal should always be plotted in order to check the linearity of the input transducer. In this case, the

coherence of both odd and even harmonics is very low, indicating that the transducer can be considered to be linear. At the system output, the coherence of the even harmonics remains low, while the coherence of the odd harmonics is very close to unity, suggesting that an odd-order nonlinear disturbance is present.

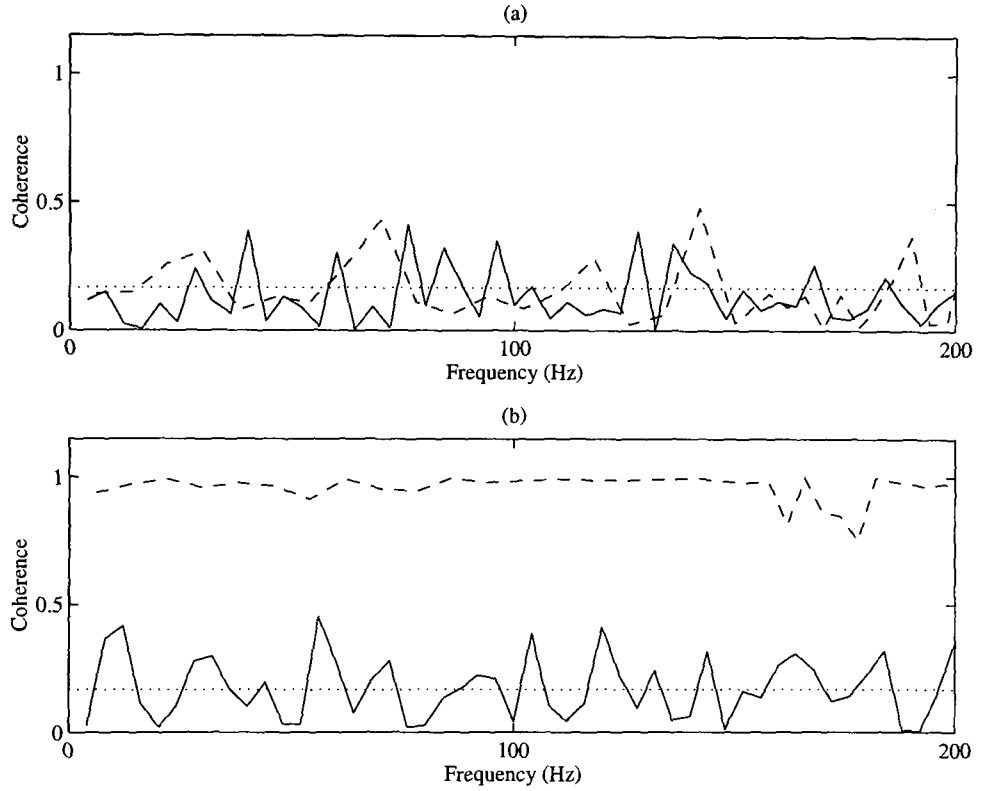

Figure 4-3. Nonlinear coherence at system (a) input and (b) output, for even (solid) and odd (dashed) excluded harmonics. The $1/M$ bound is shown dotted.

The single test has thus yielded information on the noise levels, the presence of nonlinearities and the dynamics of the system. Once the bandwidth of the system has been established then test signals can be designed which concentrate their power in the system pass-band.

If nonlinearities are detected, it is necessary to decide whether their level is significant. If their magnitude is much greater than the noise components, as seen in Figure 4-2, they are likely to be the dominant source of error in subsequent estimation. The final decision will however depend on the purpose of the identification. Small levels of nonlinearity may be acceptable if only an approximate model is required for control purposes. If an accurate model is required then steps will have to be taken to reduce or eliminate the effects of the nonlinearity, which will be discussed in Chapter 5.

4.4 No Interharmonic Distortion Signals

The sparse odd signals defined by equations (4-19) and (4-20) are made up of regularly spaced harmonics, in which only the width of the spacing between the harmonics is varied. This means that Type II contributions will always be generated at the input frequencies, since the sums and differences of different harmonics will tend to fall at the harmonics themselves.

If the spacing between the harmonics is allowed to vary freely and the sparsity of the signals is further increased, a point is eventually reached where no Type II contributions are generated at the test frequencies, for a given order of nonlinearity. This involves selecting a harmonic vector **i** so that the following relation holds, for the smallest possible maximum value of **i**

$$\mathbf{i} \notin \mathbf{s} \qquad (4-21)$$

where $\mathbf{s} = \sum_{\substack{m_1=-F \\ m_1 \neq 0}}^{F} \sum_{\substack{m_2=-F \\ m_2 \neq 0}}^{F} \cdots \sum_{\substack{m_n=-F \\ m_n \neq 0}}^{F} \sum_{p=1}^{n} i(m_p)$

Any harmonic combinations which can be grouped as a single harmonic plus pairs of equal positive and negative harmonics are excluded from the set **s**, since they correspond to Type I contributions. This design criterion generates very sparse signals, such as the following signal of 15 harmonics, which exhibits this property for a cubic nonlinearity and has an SI of 0.96.

$$\mathbf{i} = [\,1 \quad 5 \quad 13 \quad 29 \quad 49 \quad 81 \quad 119 \quad 141 \quad 207 \quad 263 \quad 359 \quad 459 \quad 543 \quad 729 \quad 775\,] \qquad (4-22)$$

Since the nonlinear distortion generated by such signals is made up of only Type I contributions, they will be termed *no interharmonic distortion* (NID) multisines. Several harmonic sets with NID properties are given in the Appendix to this chapter. The signals have the interesting property that altering their harmonic phases does not affect the nonlinear distortion at the test frequencies. This property is illustrated in Table 4-2, which shows the amplitude and phase at the output of a cubic nonlinearity for five harmonic consecutive and NID multisines, of equal input amplitude and phase.

TABLE 4-2

CUBIC OUTPUT FOR CONSECUTIVE AND NID MULTISINES

Input		Output - Consecutive		Output - NID	
Amplitude	ϕ (rad)	Amplitude	ϕ (rad)	Amplitude	ϕ (rad)
1	1	40.9	0.6	27	1
1	1	50.9	0.8	27	1
1	1	55.4	1.0	27	1
1	1	52.8	1.1	27	1
1	1	42.9	1.1	27	1

For the consecutive signal it is seen that the Type II contributions increase the overall nonlinear distortion and alter the phase. With the NID signal only Type I contributions are present, and hence the amplitude of the distortions is the same at each frequency, with the same phase as the original input harmonics. The contributions are simply a function of the number of harmonics in the input signal and their amplitudes at the input to the nonlinear element. It might be expected that minimising the number of nonlinear contributions in this way would also minimise the resulting nonlinear distortion. This is a common assumption, which will be shown to be false in Chapter 5.

From equation (4-12), the cubic nonlinear contributions at each positive test frequency ω_k of the NID signal can be expressed as

$$T_1(\omega_k) = a_3\left(3\,a^3(k) + 6\,a(k)\sum_{p=1}^{k-1} a^2(p) + 6\,a(k)\sum_{p=k+1}^{F} a^2(p)\right)$$

$$= a_3\,6\,a(k)\left(\frac{a^2(k)}{2} + \sum_{p=1}^{k-1} a^2(p) + \sum_{p=k+1}^{F} a^2(p)\right) \qquad (4-23)$$

The harmonic phases have been omitted from equation (4-23) since they are each equal to those at the input of the nonlinear element. The combination of the test frequency with a positive and negative frequency pair of another frequency is weighted by six, since there are 3! possible permutations of the three different frequencies. The combination of the

test frequency with its own positive and negative frequencies is only weighted by three, since the test frequency appears twice in this combination, which reduces the number of possible permutations.

It can be seen from equation (4-23) that the Type I contributions at the test frequencies can be considered as the multiplication of each test frequency amplitude by a weighted sum of the squared input amplitudes. Since the weighting is not equal for all amplitudes the sum of squares term will vary from frequency to frequency. However, if the number of frequencies is large then equation (4-23) can be approximated by

$$T_1(\omega_k) \approx a_3 6\, a(k) \sum_{p=1}^{F} a^2(p) \qquad (4-24)$$

and the Type I contributions can be considered to be proportional to each input amplitude multiplied by a constant, itself proportional to the total input power. This transforms the summation of linear and nonlinear paths expressed in equation (4-3) into a linear multiplication. This property can be exploited to measure the linear dynamics of systems such as the Wiener-Hammerstein model shown in Figure 4-1 (d). If the nonlinear terms are restricted to a maximum of the fourth-order, then applying a NID signal will result in an output at the test frequencies which can be expressed as

$$Y(j\omega_k) = a(k) e^{j\Phi_k} L_1(j\omega_k)\, L_2(j\omega_k) \left(1 + a_3 6 \sum_{p=1}^{F} (a(p) \mid L_1(j\omega_p) \mid)^2 \right) \qquad (4-25)$$

This will result in an estimated FRF of

$$\hat{H}(j\omega_k) = L_1(j\omega_k)\, L_2(j\omega_k) \left(1 + a_3 6 \sum_{p=1}^{F} (a(p) \mid L_1(j\omega_p) \mid)^2 \right) \qquad (4-26)$$

where the nonlinear influence has been reduced to a multiplicative gain term, which depends on the sum of the squared signal amplitudes at the output of the first linear system. While this does not minimise the complex error, defined in equation (4-2) as the mean distance between the true linear FRF and the measured FRF, it does allow the dynamics of the cascaded linear systems to be measured.

Illustrative Example. A Wiener-Hammerstein system was simulated, composed of two underdamped second-order linear systems and a static nonlinearity in which the highest term was fourth-order. The first linear system had a resonant peak at around 1,500 Hz and the second at around 3,000 Hz. A 30 harmonic NID signal was employed and the amplitude of the resulting FRF is shown in Figure 4-4 (a), along with the FRF of the two linear systems in cascade.

Figure 4-4 (b) shows the ratio of the two amplitudes, which is seen to be more or less constant except in the region of the first resonant peak. This illustrates that equation (4-25) is only an approximation and that the scaling is not strictly proportional to the amplitudes at the input of the nonlinear element. The additive Type I terms are relatively smaller where the amplitudes at the input to the nonlinear element are largest, causing a reduction in the ratio in this region. It can be seen that no such effect occurs at the resonant peak of the second linear system, which follows the nonlinearity. This error can be reduced by increasing the number of input harmonics.

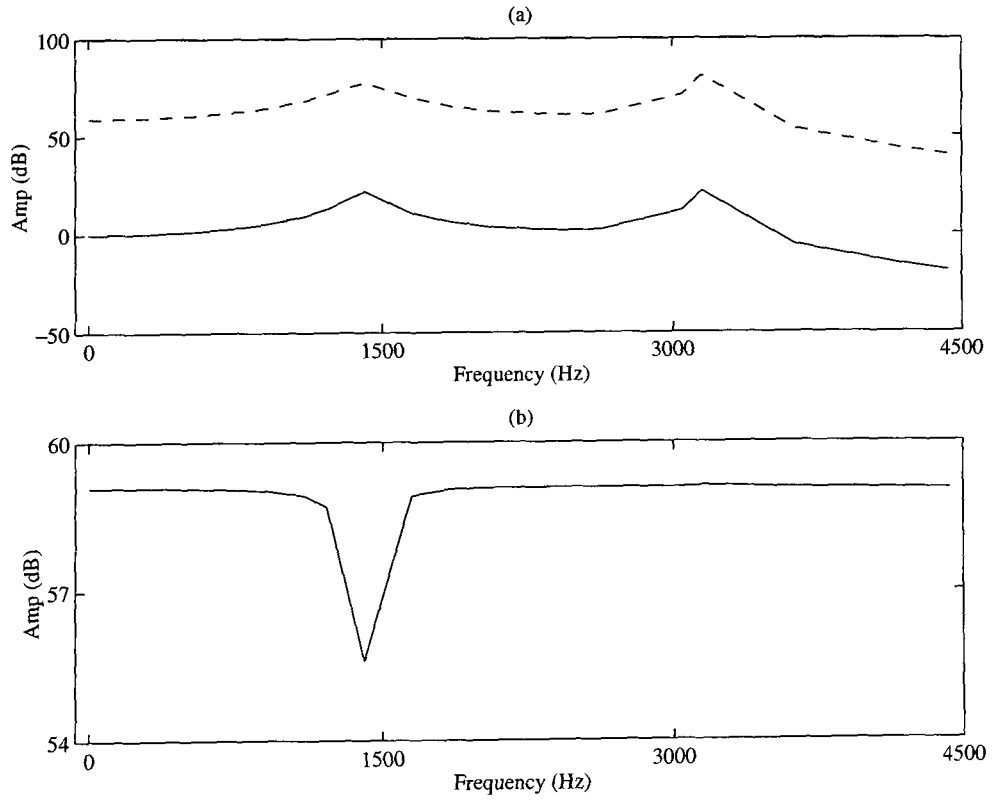

Figure 4-4. Amplitude of frequency response for Wiener-Hammerstein model. Showing (a) the amplitude of the linear systems in cascade (solid) and measured with a NID signal (dashed) and (b) the ratio of the amplitudes.

4.4.1 The Related Linear Dynamic System

Recent work conducted by Schoukens *et al.* (1998) examined the influence of power series nonlinearities on multisine signals with uniformly distributed random phases. They showed that if the number of input harmonics is large and a large number of realisations of the signals are generated, then the Type II contributions will tend to vary randomly at each output frequency. By averaging the FRF's estimated with each realisation they eliminated the random Type II influence, leaving only the Type I contributions, which are independent of phase.

Schoukens *et al.* defined the system measured in this way as the *related linear dynamic system* (RLDS) and postulated that it could be considered the best linear approximation to the overall nonlinear system

$$H_{RLDS}(j\omega_k) = H(j\omega_k) + H_b(j\omega_k) \qquad (4-27)$$

where $H_b(j\omega_k)$ is a systematic bias. For the special case of a Wiener-Hammerstein system they showed that the RLDS is equal to the underlying linear dynamics, multiplied by a scaling factor. They also showed that the RLDS corresponds to the classical results obtained on nonlinear systems using correlation techniques, with zero-mean Gaussian noise as an input.

The measurement of the same RLDS can be conducted in one experiment using a NID signal and it can be shown that the frequency response measured by the random phase signals will converge to that measured with the NID signal as the number of averages increases. This can be illustrated by examining the evolution of the complex error of the mean transfer function with the number of averages. Figure 4-5 shows the variation of the complex error for NID and random phase signals of 30 equal amplitude harmonics, applied to a cubic nonlinearity.

The analysis presented in the previous section allows some further clarification of the nature of the scaling factor introduced by the nonlinearity. The factor depends on the harmonic amplitudes of the signal applied to the nonlinear element and hence depends not only on the original input signal but also on the dynamics of the first linear system. It was also shown that the assumption of a constant multiplicative factor at each input frequency is only an approximation, which improves as the number of input frequencies is increased.

The simulation described in the previous section can be repeated to illustrate the measurement of the RLDS with random phase and NID signals. The FRF's are shown in Figure 4-6, where the response of the two linear systems in cascade is shown as a solid line and the result of a single measurement with a NID signal is shown as a dashed line. The mean FRF calculated with the random phase signals across 200 experiments is shown as a series of crosses. It can be seen that the FRF measured with the random signal is scattered around that measured with the NID signal.

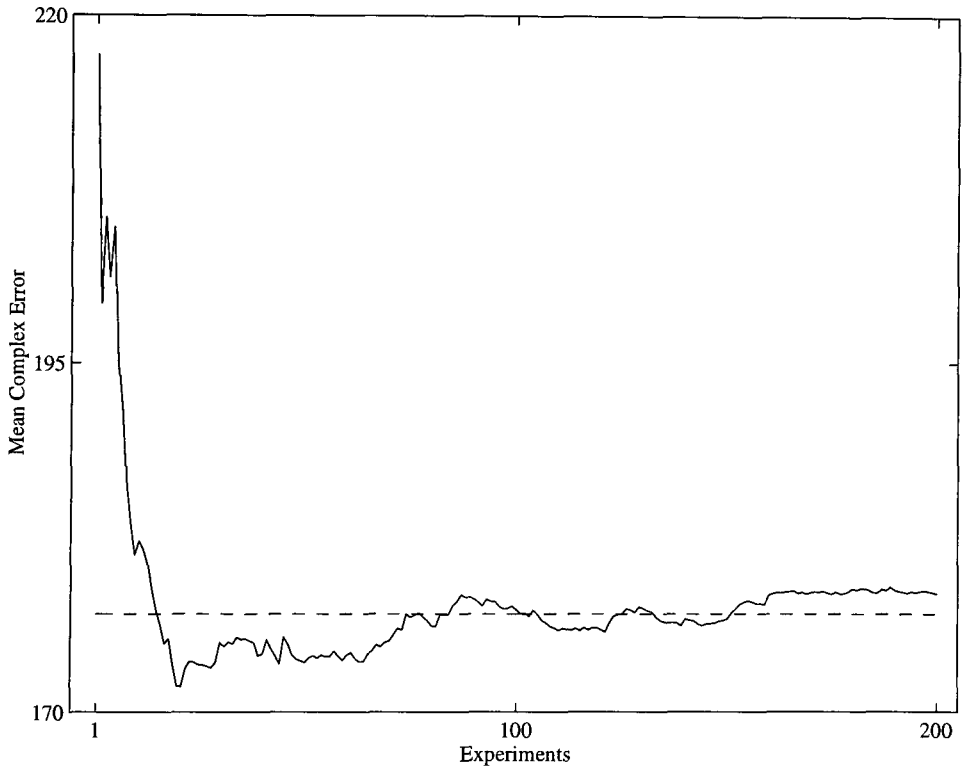

Figure 4-5. Mean complex error versus number of experiments, for a NID signal (dashed) and a random phase odd harmonic signal (solid).

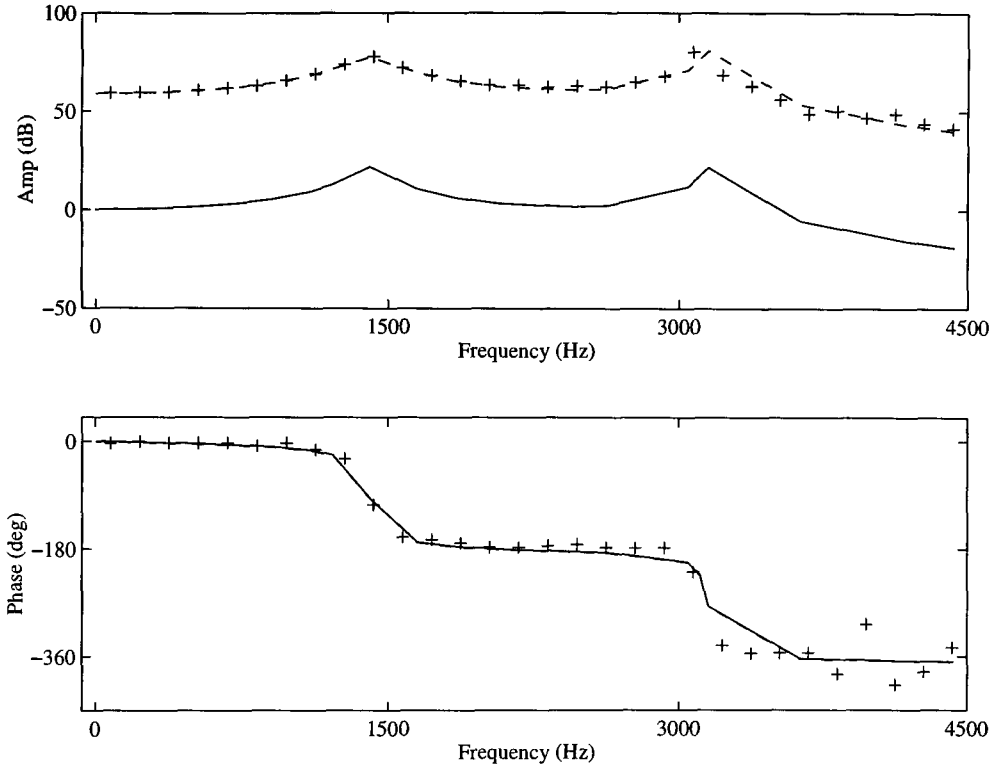

Figure 4-6. Measurement of the RLDS with a NID signal (dashed) and a random phase odd harmonic signal (crosses). The cascaded linear systems are shown as a solid line.

4.5 Conclusions

The influence of a class of nonlinear systems on multisine signals has been examined in this chapter. A novel methodology was presented for studying the frequency-domain contributions generated by static polynomial nonlinearities, which was then extended to nonlinear systems with memory.

The analysis divided the contributions generated by odd-order nonlinearities at the excitation frequencies into two types. The Type I contributions cannot be influenced by the selection of the input harmonics and depend only on the order of the nonlinearity and the number and amplitude of the harmonics at the input to the nonlinear element. By contrast, the number of Type II contributions which fall at the input frequencies depends entirely on the specific harmonics included in the signal.

The new analysis method provides a far better insight into the influence of the nonlinearity than the conventional approach, which is to simply divide the nonlinear contributions into nth harmonics and intermodulation harmonics.

A class of odd harmonic multisines was defined which include spaces in their spectra for the detection of nonlinear harmonics, which were termed sparse odd signals. These can be employed with the nonlinear coherence function to detect the presence of periodic nonlinear contributions at the excluded frequencies. This quantity will decrease as $1/M$ if only noise is present at a given frequency, which provides a useful bound in assessing the periodicity of any output components.

The design methodology was then used to design sparse odd signals which do not generate any Type II contributions at the input frequencies, for a given order of nonlinearity. Such signals were termed *no interharmonic distortion* multisines and it was shown that for a Wiener-Hammerstein system they can be used to measure the underlying linear dynamics multiplied by a gain factor. This gain factor can be considered as constant at each input frequency, if the number of frequencies is high, and depends on the harmonic amplitudes of the input signal and the dynamics of the first linear system.

This was shown to be closely related to work conducted by Schoukens *et al.* (1998) using multisines with uniform random phases. By averaging the frequency response measured with different random phased signals they arrived at a measurement which they defined as the *related linear dynamic system*, which can be considered as the best linear approximation of the overall nonlinear system. This approximation is dependent on the total power of the input signal and a given result will be specific to that signal and the input amplitude used. Application of a NID signal removes the need to average across many experiments, with the obvious restriction that a multisine with only the specified harmonics must be imposed on the system.

4.6 Appendix - Harmonic Vectors of NID Multisines

The following harmonic vectors have NID properties for nonlinearities up to and including the fourth order.

```
i=[   1     5    13    29    49    81 ...
    119   141   207   263   359   459 ...
    543   729   775   909  1097  1213 ...
   1405  1649  1853  2077  2461  2635 ...
   3047  3111  3151  3631  4177  4431 ]
```

```
i=[    1   151   305   461   619   779 ...
     941  1105  1271  1439  1609  1781 ...
    1955  2131  2309  2489  2675  2857 ...
    3045  3229  3421  3611  3805  4003 ...
    4203  4437  4641  4891  5265  5547 ]
```

Chapter V

Minimising the Nonlinear Distortion

Abstract — *This chapter examines the effects of nonlinear distortions on the frequency-domain estimation of linear models using multisine test signals. The aim is to minimise the distortion introduced by the nonlinearity, for a given input power constraint. A number of different multisine signals are compared for this purpose, with zero, random and low CF harmonic phases. The results are compared with those of other authors in this field. The elimination of the nonlinear effect is then addressed, using a two-level multisine test. An overall identification scheme is outlined for obtaining high quality linear models of systems suffering nonlinear distortions.*

5.1 Introduction

It is often desirable to obtain a linear model of a system suffering some nonlinear distortions. This can simplify both the analysis of the system itself and, if required, the design of an appropriate controller. The utility of the simplified linear model must then be established in practice. This chapter examines the design of test signals for this purpose. It is assumed that the systems under study have underlying linear dynamics which dominate the system response at low input signal amplitudes. The influence of the nonlinearity on the test signals is to be minimised, for a given input power constraint. The aim is thus to isolate the linear component of the system response and not, as in Chapter 4, to obtain the best linear approximation of the overall nonlinear system. The linear dynamics identified in this way will be independent of the input signal and its amplitude.

Important insights are offered into the effects of nonlinearities on different multisines, which emphasise the importance of minimising their CF. If a significant odd-order nonlinearity is present then steps must be taken to eliminate its effects and a compensation technique will have to be employed. A technique is considered which will eliminate a given order of nonlinear distortion for nonlinearities which can be described by a series expansion, such as static power series, Volterra and Wiener models. This also allows the influence of the nonlinearity on the measured data to be quantified.

5.2 Comparing Multisines

The analysis presented in Chapter 4 showed that it is not possible to eliminate the effect of odd nonlinearities simply by adjusting the harmonic content of the test signal, since Type I contributions will always be generated. A class of NID multisines was defined, which generate only Type I contributions at the test frequencies, for a given order of nonlinearity. These allow the measurement of the best linear approximation to the nonlinear system, termed the related linear dynamic system.

However, this is not the same as measuring the underlying linear dynamics, since the gain of the RLDS is dependent on the input signal amplitude, which means that it can be very different from that of the linear system. The aim of this chapter is to obtain a frequency response which lies as close as possible to that of the underlying linear dynamics, for a given nonlinearity and input signal power. If this is to be achieved without applying further processing, then a signal must be selected which minimises the complex error defined in equation (4-2).

A comparison will thus be made of the distortion generated by a range of different multisines of equal power, while varying the number of harmonics. The stipulation of equal power ensures that the influence of stochastic effects will be the same for all the signals. The results are presented for a cubic nonlinearity but can be generalised to higher power odd terms, in accordance with the analysis described in Chapter 4.

Multisines of increasing sparsity will be compared, composed of consecutive, odd, prime, odd-odd and NID harmonics. Prime harmonic signals were first proposed by Rees (1976) and are composed of only those harmonics which are prime multiples of the fundamental, with the fundamental and the second harmonic excluded. Rees argued that such signals would be less affected by nonlinear distortions, since the nonlinear contributions which were multiples of the input frequencies would not fall at any of those frequencies. The analysis presented in Chapter 4 shows that Type I contributions will naturally fall at the prime harmonics and that Type II contributions made up of sums of *different* input harmonics will also fall at the these harmonics. For example, the sums of prime harmonics $\delta(\omega-(13\omega_0+3\omega_0-11\omega_0))$ and $\delta(\omega-(7\omega_0+3\omega_0-5\omega_0))$ will both generate contributions at the prime harmonic $5\omega_0$.

An initial comparison was made between multisines of five harmonics, in order to investigate the generation of Type I and Type II contributions at the test frequencies. The NID signal was composed of the first five harmonics of the vector given as equation (4-22) and the prime harmonics were as follows

$$\mathbf{i}=[3 \quad 5 \quad 7 \quad 11 \quad 13] \tag{5-1}$$

The number of disturbing contributions generated by the cubic nonlinearity was found by considering the multiple summation of impulses in equation (4-10) and counting only those contributions which fell at the positive input harmonics.

$$C=\text{count}\left[\sum_{p=-F}^{F}\sum_{m=-F}^{F}\sum_{k=-F}^{F}[i(p)+i(m)+i(k)] \mod \mathbf{i}\right] \tag{5-2}$$

The number of Type I contributions is equal in each case and the consecutive, odd, prime and odd-odd harmonic signals also generate Type II contributions at the test frequencies, a comparison of which is made in Table 5-1.

TABLE 5-1

NONLINEAR CONTRIBUTIONS FOR VARIOUS MULTISINES

Test	Type I	Type II				
Freq.		Consec.	Odd	Prime	Odd-Odd	NID
1	27	36	48	27	18	0
2	27	36	46	27	27	0
3	27	34	42	18	30	0
4	27	30	36	18	27	0
5	27	24	28	18	18	0

The largest number of Type II contributions is generated by the odd harmonics, followed by the consecutive, the odd-odd harmonics and then the prime. This pattern is repeated for signals with larger numbers of harmonics. Thus it might be concluded that the NID signal is to be preferred, since the minimum possible number of contributions fall at the test frequencies with this signal. However, the overall error introduced by the sum of the Type I and Type II contributions will depend on their relative phases, as will now be shown.

The five harmonic signals were used to illustrate the number of contributions which fall at each input harmonic but they are unlikely to be used in practice. The error generated by the cubic nonlinearity will now be examined for more practical signals composed of ten, twenty and thirty harmonics. For the ten harmonic signal, the double-sided spectral amplitude of each harmonic was set equal to one. The other signals were then scaled to ensure that the total signal power remained constant and the harmonic phases were varied. The simulated system was simply a static cubic nonlinearity with the coefficient a_3 set to one.

The complex error was calculated using equation (4-2) and the results for signals with zero phases are shown in Table 5-2. This is the worst-case scenario, since the Type I and Type II contributions are all in-phase and hence additive. The largest errors are obtained with the odd harmonic signal and the smallest with the NID harmonics, as would be

expected from Table 5-1. The complex error is seen to increase with the number of harmonics, which is to be expected, since a larger number of in-phase contributions are generated. The use of zero-phase signals for practical testing is not recommended.

TABLE 5-2

COMPLEX ERROR FOR ZERO PHASE MULTISINES

F	Consec.	Odd	Prime	Odd-Odd	NID
10	249	267	178.2	201	57
20	514.5	533.5	318.9	400.5	58.5
30	780.78	800.11	450.2	600.33	59

The relative phases were then set to uniformly distributed random values, in the interval $[0,2\pi]$, and Table 5-3 lists the complex error of the mean transfer function over 200 experiments. The error obtained with the NID signals is constant in each case, since it is unaffected by the varying phases of the Type II contributions. It can be seen that the errors obtained with each of the other signals are close to those obtained with the NID signals. This confirms the conclusion drawn in Chapter 4, that the influence of the Type II contributions is approaching zero across the 200 experiments.

Another significant point is that the error is increasing with the number of harmonics, but only very slightly. This error is almost entirely a function of the Type I contributions, which are increasing in number but decreasing in amplitude as the number of harmonics is increased, due to the equal power constraint.

TABLE 5-3

COMPLEX ERROR FOR RANDOM PHASE MULTISINES

F	Consec.	Odd	Prime	Odd-Odd	NID
10	57.1	57.2	58.4	56.2	57
20	59.1	58.6	59.0	59.0	58.5
30	59.6	59.5	59.5	59.6	59

The effect of using low CF phases was studied by compressing the signals using the $L\infty$ method discussed in Chapter 3, and the resulting CF are shown in Table 5-4. The CF of the NID signals is poor compared to the other multisines, as would be expected for signals with such sparse spectra.

Using the low CF signals in the nonlinear simulation produced some interesting results, which are presented in Table 5-5. It can be seen that the resultant error with the odd, consecutive, prime and odd-odd harmonics is now smaller than that with the NID harmonics. The results for the prime and odd-odd harmonics are quite similar, which suggests that the selection of prime harmonics is not a key factor in influencing the nonlinear distortion.

TABLE 5-4

MULTISINE CREST FACTORS, $L\infty$ METHOD

F	Consec.	Odd	Prime	Odd-Odd	NID
10	1.47	1.44	1.54	1.61	2.38
20	1.42	1.42	1.59	1.59	2.99
30	1.41	1.36	1.61	1.58	3.05

TABLE 5-5

COMPLEX ERROR FOR LOW CF MULTISINES

F	Consec.	Odd	Prime	Odd-Odd	NID
10	31.1	30.8	32.2	35.2	57
20	29.9	30.1	34.4	34.2	58.5
30	29.8	28.2	34.3	33.5	59

This difference in complex error between the low CF and random phase signals must be attributed to the difference in number and effect of the Type II contributions. Rather than adding to the nonlinear distortion, the Type II contributions appear to lessen it. This suggests that the phases of low CF signals result in Type II contributions which are out of phase with the Type I contributions, so as to reduce the overall nonlinear distortion.

Comparing the different low CF signals, it can be seen that the Type II contributions generated by the consecutive and odd harmonics have a more significant anti-phase effect on the Type I contributions, than those generated by the prime or odd-odd harmonics. This presents an interesting analogy between the effects of the nonlinearity on signals of different CF, in the time and frequency domains. The signals with a lower CF have a smaller maximum time-domain amplitude, given the equal power constraint, and the effect of the nonlinearity will be reduced. In the frequency domain, the amplitude of the harmonics will be the same, hence the reduction in the nonlinear effect must be attributed to the number and phase of the Type II contributions, which act in anti-phase to the Type I contributions.

This suggests that a signal which is amenable to CF minimisation will be one which generates many interacting Type II contributions. Such signals have dense, evenly spaced spectra and under these conditions the Type II contributions significantly reduce the overall nonlinear contribution. The phases of low CF signals thus correspond to Type II contributions which are out of phase with the Type I contributions.

5.2.1 Practical Results

The simulation results were practically verified using consecutive, odd and prime harmonic multisines. The effects of even- and odd-power nonlinearities were assessed by adding a controlled nonlinear distortion to the output of a first-order active low-pass filter, which had a break-point at 100 Hz. A linear model of the filter was first identified using the frequency-domain estimator described in Chapter 2. Signal averaging was used to achieve SNR's of greater than 70 dB at the input and output and hence a high quality linear model. The nonlinear distortion was then generated by a simple circuit made up of analogue multipliers and a voltage divider. The influence of input and output saturation on the signals was also investigated.

A reconstruction filter was used to remove the high frequency butterflies at the output of the ZOH, which would otherwise interact with the test frequencies and introduce additional distortion, due to the presence of a nonlinearity. Multisines of twenty harmonics were employed, which were scaled to have equal total power. The signals were once again averaged to obtain SNR's of better than 70 dB and ensure that the stochastic errors were very small compared to those introduced by the nonlinearities. Estimates of the FRF's were calculated using the H_{EV} estimator defined in equation (2-37) and first-order s-domain models were also estimated in the frequency domain.

The complex error E was calculated between the known linear frequency response and the estimated FRF's and is listed in Table 5-6. The influence of the second-order nonlinearity on the consecutive multisine is clear. In the case of the cubic nonlinearity and the input and output saturation, the error is almost equal for both the consecutive and odd harmonic multisines, while it is a little higher for the prime harmonic signal. A more detailed comparison of the signals is presented in the following sections.

Second-Order Nonlinearity. The magnitude and phase plots of the actual and estimated transfer functions with a second-order distortion are shown in Figures 5-1, 5-2 and 5-3. The FRF estimates, shown as crosses, indicate that the odd and prime harmonic signals are immune to second-order distortions, while the consecutive signal is badly affected. The frequency response of the estimated parametric model $\hat{H}(s)$, shown as a dashed line, is good for all signals. This result should be treated with some caution, since this might not be the case with higher-order systems.

TABLE 5-6

COMPLEX ERROR FOR MULTISINES IN PRACTICAL TESTS

Signal	Second-order	Third-order	I/P saturation	O/P saturation
Consecutive	0.095	0.041	0.034	0.028
Odd	0.004	0.039	0.029	0.026
Prime	0.004	0.056	0.069	0.043

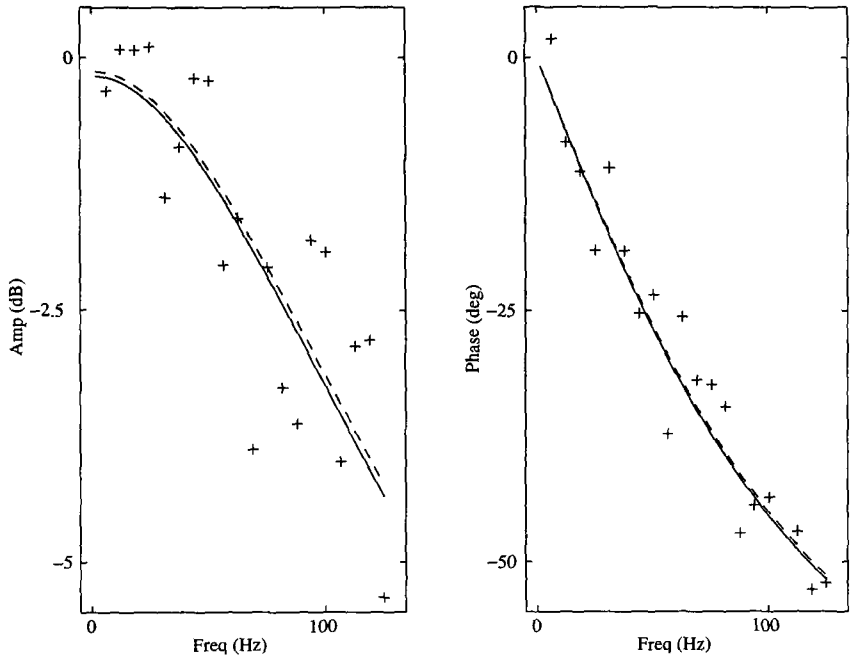

Figure 5-1. Actual and estimated transfer functions, second-order nonlinearity, consecutive harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed, and $H_{EV}(j\omega)$ - crosses.

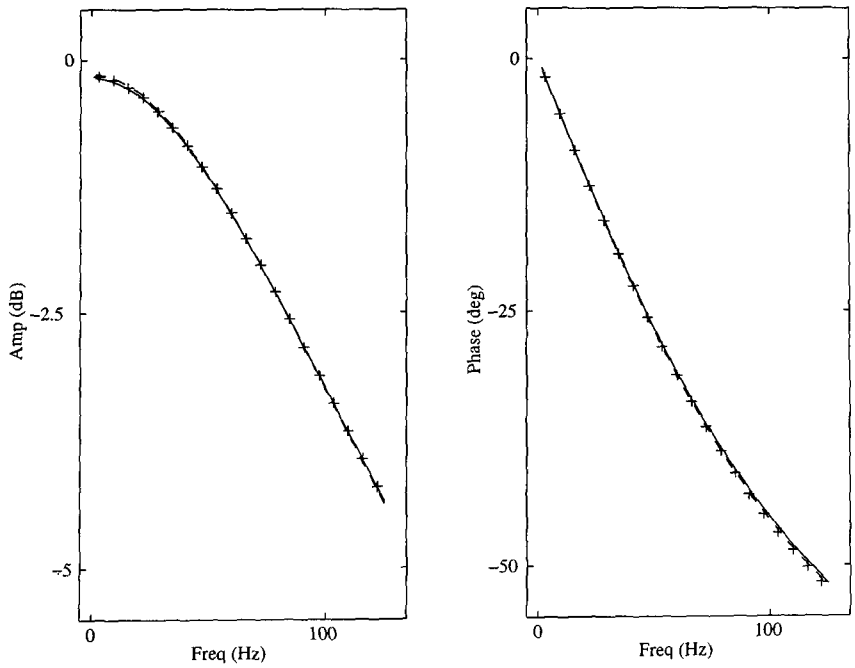

Figure 5-2. Actual and estimated transfer functions, second-order nonlinearity, odd harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$- dashed, and $H_{EV}(j\omega)$ - crosses.

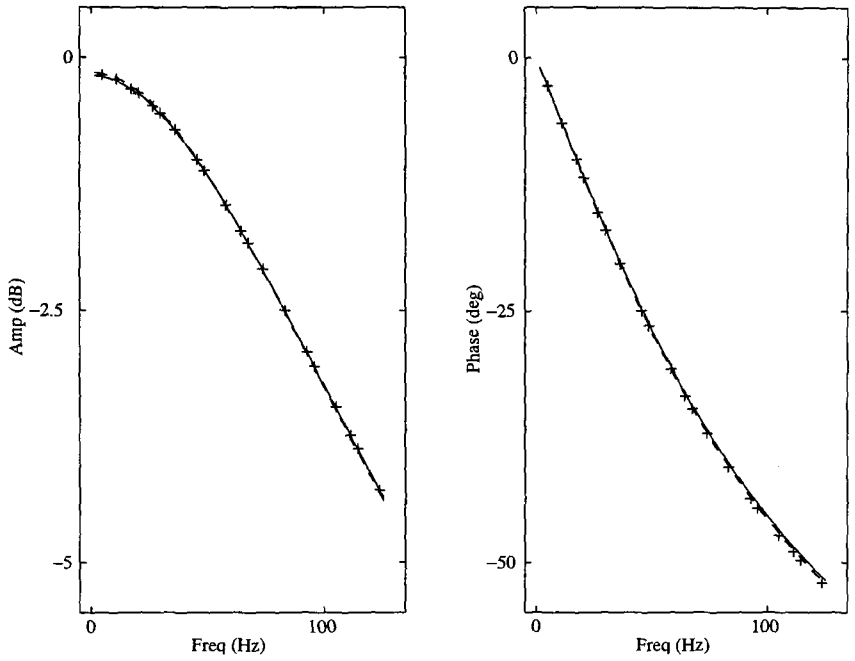

Figure 5-3. Actual and estimated transfer functions, second-order nonlinearity, prime harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed, and $H_{EV}(j\omega)$- crosses.

Third-Order Nonlinearity. Results for the third-order distortion are presented in Figures 5-4, 5-5 and 5-6. They show the multisines to be equally affected, with no advantage to using the prime signal, since the nonlinearity introduces a similar distortion for each of the signals. There is a significant bias in the phase, whereas the amplitude measurements appear to be scattered around the true linear response.

Output Saturation. The same conclusions can be drawn from the results for output saturation, shown in Figures 5-7, 5-8 and 5-9. As would be expected, the most significant bias is in the amplitude in this case, with an apparent reduction in the dc gain of the system. Similar results were obtained for saturation at the system input.

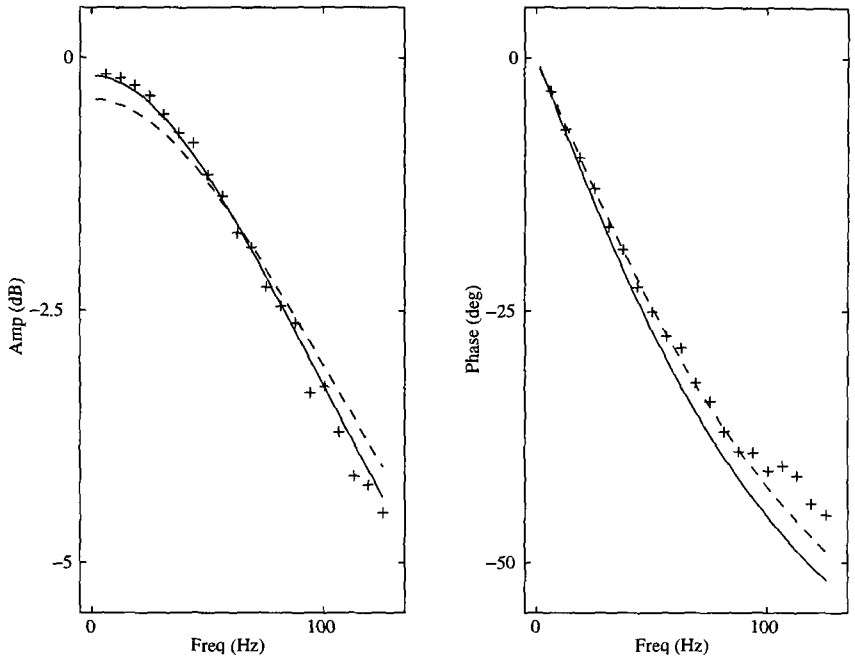

Figure 5-4. Actual and estimated transfer functions, third-order nonlinearity, consecutive harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed and $H_{EV}(j\omega)$ - crosses.

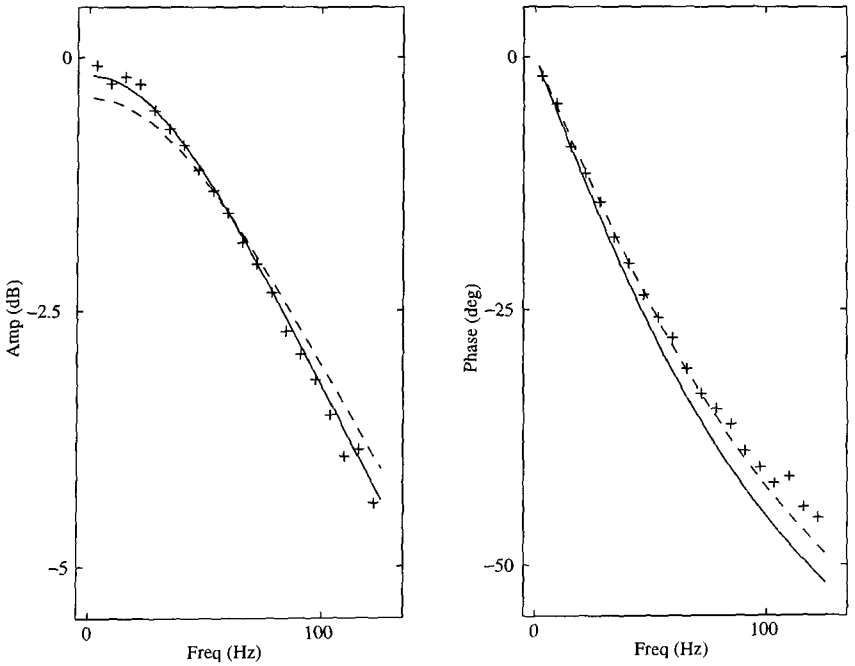

Figure 5-5. Actual and estimated transfer functions, third-order nonlinearity, odd harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed, and $H_{EV}(j\omega)$ - crosses.

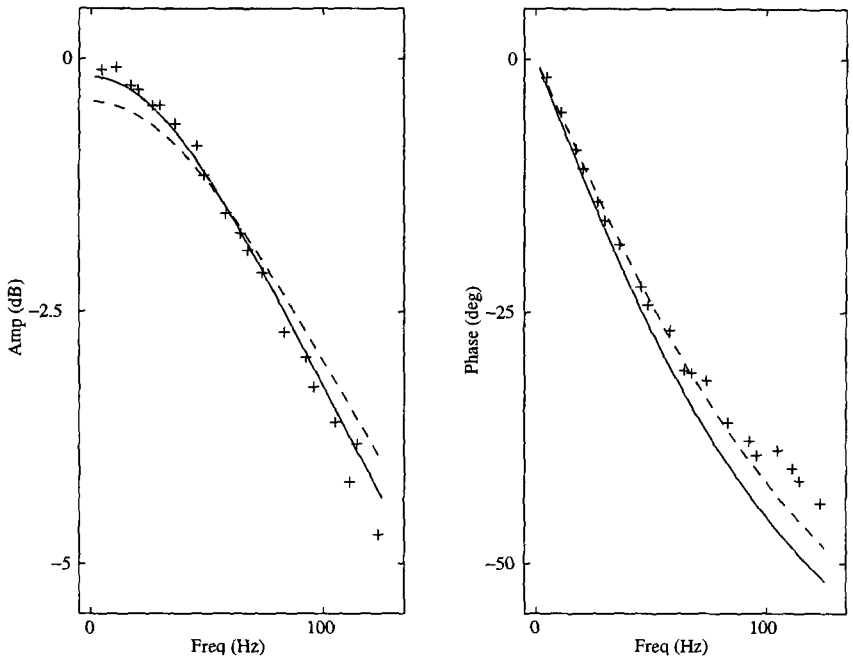

Figure 5-6. Actual and estimated transfer functions, third-order nonlinearity, prime harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed, and $H_{EV}(j\omega)$ - crosses.

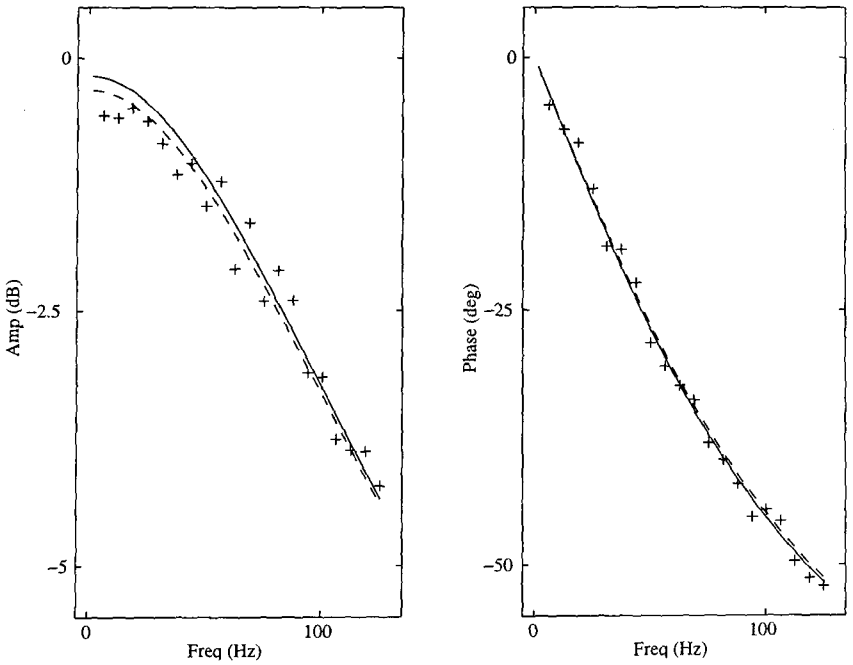

Figure 5-7. Actual and estimated transfer functions, output saturation, consecutive harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed, and $H_{EV}(j\omega)$ - crosses.

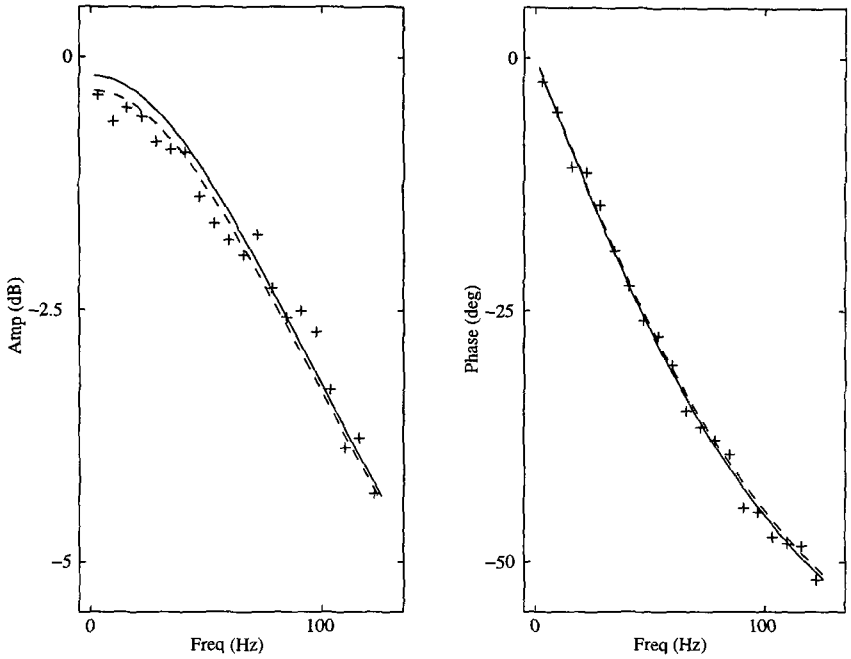

Figure 5-8. Actual and estimated transfer functions, output saturation, odd harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed and $H_{EV}(j\omega)$ - crosses.

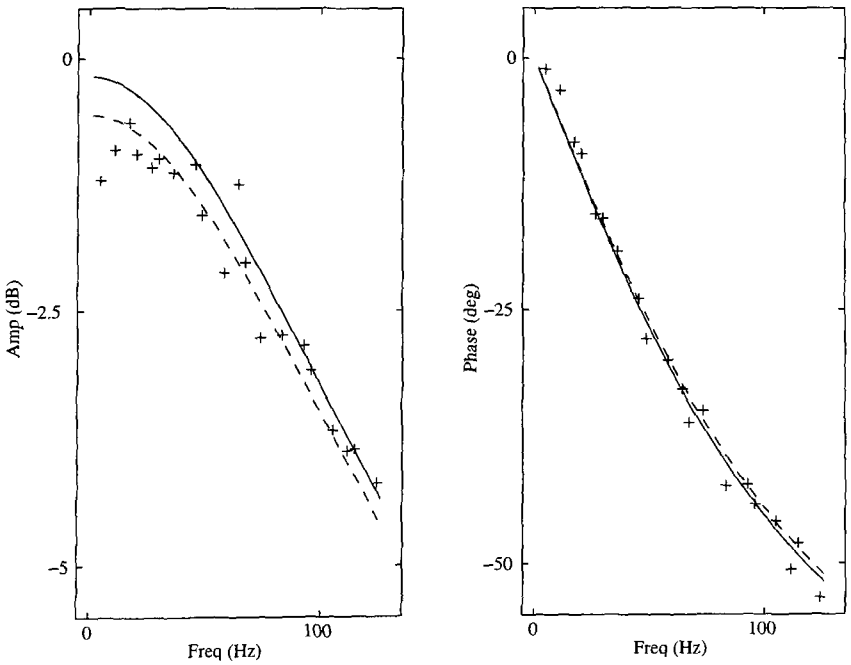

Figure 5-9. Actual and estimated transfer functions, output saturation, prime harmonic multisine. $H(s)$ - solid, $\hat{H}(s)$ - dashed and $H_{EV}(j\omega)$ - crosses.

5.2.2 Discussion

It is seen that a systematic error is inherent in any system suffering from odd-order nonlinear distortions, since nonlinear components will be generated at the test frequencies whatever the frequency content of the signal. The dominant variable is thus the CF of the signal. The relationship between CF and the complex error E is illustrated in Figure 5-10 for odd, consecutive and NID harmonic multisines of equal power. This result was obtained by randomly varying the phases of the signals in 10,000 trials and recording the CF and respective complex error. The results for the consecutive and odd signals lie over each other and are only distinguishable at higher CF.

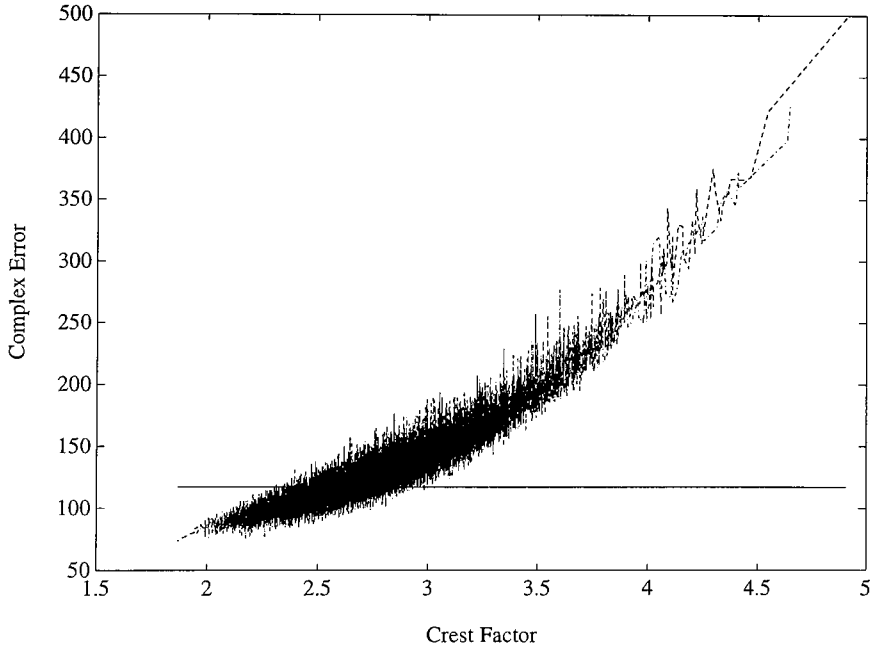

Figure 5-10. Complex Error (E) versus CF for 20 harmonic multisines, NID (solid), odd (dashed) and consecutive (dash-dot).

As would be expected, the error is constant with the NID signal. For the odd and consecutive signals it is seen that the error can vary considerably between signals with close values of CF. This suggests that the CF is not the only variable affecting the complex error, but the overall trend is clear and indicates that the complex error is proportional to the square of the CF of the signals.

It is seen that the error obtained with the odd and consecutive signals is less than that with the NID once the CF falls below around 2.3. Such a CF is easily achieved in practice,

even using the simple formula proposed by Schroeder (1970). However, if the phases are set to random values then a CF worse than 2.2 will often be obtained. In this study, mean CF of around 2.7 were obtained for odd and consecutive signals using uniformly distributed random phases over 10,000 trials.

The prime multisine showed no greater immunity to odd-order nonlinear distortions than a consecutive, odd or odd-odd harmonic signal. While it was shown in Table 5-1 that the number of Type II contributions which fall at the test frequencies is indeed reduced with the prime harmonics this did not prove to be the key factor in reducing the nonlinear effect. When low CF signals were employed the exact opposite was shown to be the case, in that the greater the number of Type II contributions the greater the reduction in nonlinear distortion.

The same argument applies to the NID signals, which have the smallest possible number of contributions falling at the test frequencies but the largest error in the case of low CF signals. This runs contrary to the common assumption that minimising the number of contributions will minimise the magnitude of the nonlinear distortion. This assumption was made by Suki and Lutchen (1992), who designed signals similar to NID multisines and used them to measure the input impedance of the respiratory systems of dogs. They compared the performance of their *no-sum, no-difference* (NSND) signals with that of a consecutive multisine and single sines.

They began by designing signals, composed of both even and odd harmonics, which did not generate any second-order contributions at the test frequencies. They called these NSND 2 signals and the harmonic vector listed in their paper is shown in equation (5-3).

$$\mathbf{i} = [\,3 \quad 7 \quad 11 \quad 17 \quad 23 \quad 29 \quad 37 \quad 41 \qquad 61\,] \tag{5-3}$$

It is clear from the discussion in Chapter 4 that immunity to even-order nonlinearities is guaranteed simply by using an odd harmonic signal, making it unnecessary to design harmonic vectors such as that given in equation (5-3). They went on to design a signal with the same properties for a fourth-order nonlinearity, termed NSND 4, while acknowledging that some nonlinear contributions would be generated at the test frequencies by any third-order term.

$$\mathbf{i} = [\,5 \quad 11 \quad 19 \quad 31 \quad 59 \quad 103 \quad 163\,] \tag{5-4}$$

They simulated a static fifth-order nonlinear system, in order to assess the performance of the signals, and made a comparison between signals of equal maximum time-domain amplitude, with random harmonic phases. It was concluded that the NSND signals generated less nonlinear distortion. The problem with this approach is that signals of unequal power and an unequal number of harmonics were compared. The consecutive signal had thirty harmonics, while the NSND 2 signal had the eleven and the NSND 4 only seven. It is therefore unsurprising that the consecutive signal generated the largest errors, since it was the only signal to be affected by the even-order nonlinear terms and also had the most harmonics.

The influence of minimising the signal CF was also not considered and while Suki and Lutchen noted the significant effect of varying the harmonic phases on the resultant error they did not pursue this line of investigation any further. A more meaningful comparison would have been to consider signals of equal power and an equal number of harmonics, with low CF phases. The simulation can be repeated under these conditions, using the nonlinear system employed by Suki and Lutchen, which is given in equation (5-5). The signals used and the results obtained are listed in Table 5-7, which shows that the odd multisine generated the smallest error.

$$y(t) = x(t) + x(t)^2 + 0.5\,x(t)^3 + 0.3\,x(t)^4 + 0.2\,x(t)^5 \qquad (5-5)$$

Suki and Lutchen also drew attention to the smoothness of the FRF estimated with the NSND signals, compared to the scattered estimates obtained with the consecutive signal. This point concurs with the analysis of the differing effects of Type I and Type II contributions made in Chapter 4. The differing nature of the error is illustrated in Figure 5-11, which shows the true linear transfer function of the simulated system and the estimated FRF's using the consecutive, odd and NSND 4 harmonics.

The FRF's estimated with the consecutive and odd harmonic signals show considerable scatter in the phase, while the phases estimated with the NSND 4 signal are much closer to the true value. However, the amplitude estimated with the NSND 4 signal shows a much greater bias than either the consecutive or odd harmonic signals, which explains why the overall complex error is considerably larger. This can be explained by the predominant influence of the Type I contributions on the NSND signal, which are generated by the odd-order nonlinear terms and introduce an in-phase bias of equal amplitude at each test point.

TABLE 5-7

COMPARISON OF CONSECUTIVE, ODD AND NSND HARMONIC SIGNALS

Signal	Harmonics	CF	E
Consecutive	1, 2, 3, 4, 5, 6, 7	1.47	9.4
Odd	1, 3, 5, 7, 9, 11, 13	1.42	8.3
NSND 2	3, 7, 11, 17, 23, 29, 37	1.69	12.2
NSND 4	5, 11, 19, 31, 59, 103, 163	2.36	30.1

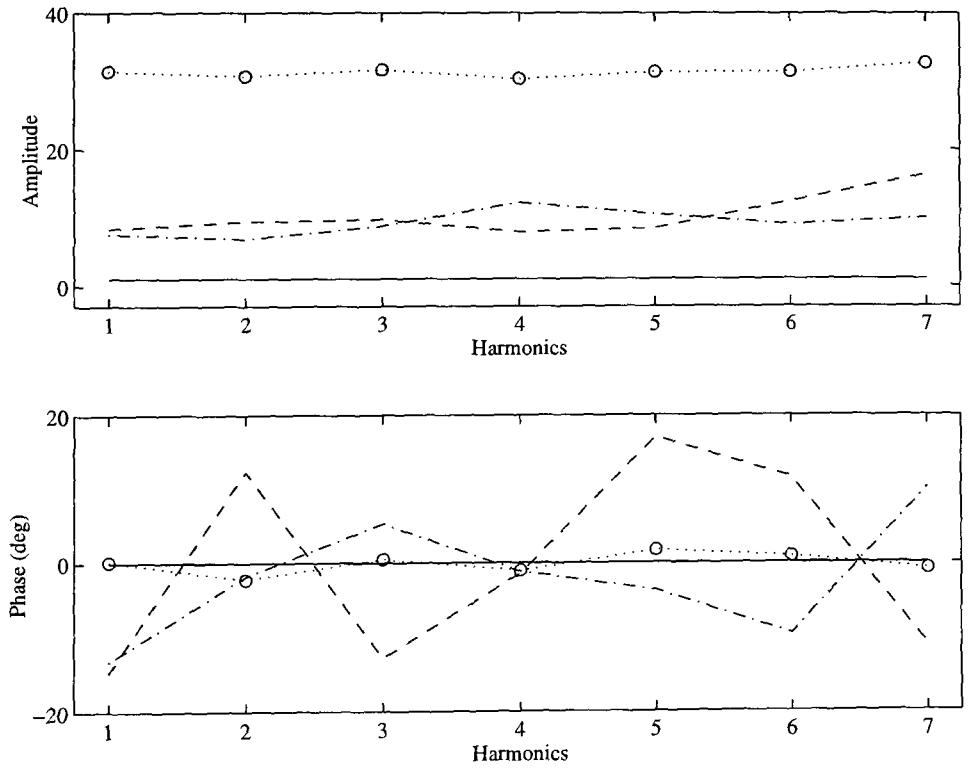

Figure 5-11. True FRF (solid) and FRF estimated with consecutive (dashed), odd (dash-dot) and NSND 4 (dotted-circle) signals.

A more recent paper by Lutchen *et al.* (1993) addressed the selection of phases in the design of low CF waveforms for ventilating human patients. Unfortunately, the superiority of the NSND signals was assumed to have been proven and signals of unequal power and harmonics were once again compared in the illustrative simulations. The NSND signal was then compared to a step input in the practical study.

It was shown earlier in this section that NID-type signals, while exhibiting some useful properties, do not offer any benefit in the *reduction* of nonlinear distortion when estimating linear models. Such signals do find application in the measurement of the RLDS, which was discussed in Chapter 4, and in the measurement of nonlinear Volterra kernels, which will be discussed in Chapter 6. The signals designed by Suki and Lutchen do possess NID properties and are suitable for the use described in Chapter 4 but not for their stated aim of minimising the nonlinear distortion at the test frequencies.

It can be generally concluded that there is no benefit in using multisines of increasing sparsity, if the main aim is to reduce the effect of odd-order nonlinearities at the test frequencies. The error can be reduced but not eliminated by minimising the signal CF. If no further processing is to be applied, the only way to further reduce the nonlinear influence is to reduce the test signal amplitude. A lower limit usually exists on this value, in order to ensure sufficiently high accuracy in the face of stochastic errors. In these circumstances, the best signal is an odd harmonic multisine, since it is immune to even-order nonlinear distortions and amenable to CF minimisation.

5.3 Compensating for Nonlinearity

If the influence of an odd nonlinearity is to be completely eliminated a compensation technique will have to be employed. The following method can be applied under the assumption that the nonlinearity can be modelled by a single valued static nonlinearity, or a functional series of the Volterra or Wiener type. Since a wide range of nonlinear systems can be modelled in this way, this is not an over-restrictive constraint.

5.3.1 Theory

A technique to eliminate a cubic nonlinear distortion was proposed by Barker and Davy (1975), using inverse repeat MLBS input signals. The method involved exciting the system with the same sequence at two different amplitudes and combining the results to eliminate the effect of the nonlinear term. A similar approach was suggested by Schoukens *et al.* (1993b) and employed with single sine tests. A modified form of these methods is proposed, employing an odd harmonic multisine as a test signal. The technique can be applied to any odd-order nonlinearity but the most natural approach is to compensate for lower order terms, in order to ensure that the linear estimate is unaffected by nonlinear distortions up to a given degree.

Consider $U_1(j\omega_k)$ and $U_2(j\omega_k)$ to be the measured input spectra of an odd multisine, differing only in amplitude, and $Y_1(j\omega_k)$ and $Y_2(j\omega_k)$ the measured responses to each input. The output signals will contain two components

$$Y_r(j\omega_k) = Y_{l,r}(j\omega_k) + Y_{nl,r}(j\omega_k) \qquad r = 1,2 \tag{5-6}$$

where $Y_{l,r}(j\omega_k)$ and $Y_{nl,r}(j\omega_k)$ are the unknown linear and nonlinear responses. The FRF's estimated directly with these signals will differ, due to the influence of the nonlinearity.

$$\frac{Y_1(j\omega_k)}{U_1(j\omega_k)} \neq \frac{Y_2(j\omega_k)}{U_2(j\omega_k)} \tag{5-7}$$

In order to eliminate the influence of a nonlinearity of order n, scaling factors must be defined which are proportional to the input signal raised to that power. Equation (4-10) can be used to calculate the output of a nonlinear term of that power, with the coefficient a_n set to one, when the input signal is applied at each amplitude. The outputs at the test frequencies are then selected

$$S_{n,r}(j\omega_k) = Y_{n,r}(j\omega_k) \qquad r = 1,2 \tag{5-8}$$

and used to scale the input and output Fourier coefficients.

$$U_s(j\omega_k) = U_1(j\omega_k)\, S_{n,2}(j\omega_k) - U_2(j\omega_k)\, S_{n,1}(j\omega_k) \tag{5-9}$$

$$Y_s(j\omega_k) = Y_1(j\omega_k)\, S_{n,2}(j\omega_k) - Y_2(j\omega_k)\, S_{n,1}(j\omega_k) \tag{5-10}$$

Substituting equation (5-6) into equation (5-10) results in

$$Y_s(j\omega_k) = Y_{l,1}(j\omega_k)\, S_{n,2}(j\omega_k) - Y_{l,2}(j\omega_k)\, S_{n,1}(j\omega_k) + [Y_{nl,1}(j\omega_k)\, S_{n,2}(j\omega_k) - Y_{nl,2}(j\omega_k)\, S_{n,1}(j\omega_k)] \tag{5-11}$$

The grouped nonlinear terms will be eliminated if the following relation holds

$$\frac{S_{n,1}(j\omega_k)}{S_{n,2}(j\omega_k)} = \frac{Y_{nl,1}(j\omega_k)}{Y_{nl,2}(j\omega_k)} \qquad (5-12)$$

Since the elimination of a given order of distortion depends only on the ratio of these quantities, the precise form of the nonlinearity is unimportant. It is sufficient that its output may be described, at least in part, by a nonlinear term of order n. For a nonlinear system with memory, described by equation (4-14), the memory term will be a complex scaling factor common to both the numerator and denominator of the right hand side of equation (5-12) and the nonlinear term will be eliminated in the same way as a static nonlinear contribution.

5.3.2 Practical Results

The technique will be illustrated using the practical system described in Section 5.2.1, with a controlled third-order nonlinearity added to the output of the linear system. The level of the nonlinearity was increased for this example, in order to generate more nonlinear distortion at the test frequencies. Measurements were made at three input amplitudes, of 0.3, 1 and 2 Volts, using a 20 odd-harmonic signal. The measured signals were averaged across six periods and scaling factors were calculated using equation (4-10) as

$$Y_{3,r}(j\omega) = \sum_{\substack{m_1=-F \\ m_1\neq 0}}^{F} \sum_{\substack{m_2=-F \\ m_2\neq 0}}^{F} \sum_{\substack{m_3=-F \\ m_3\neq 0}}^{F} \prod_{p=1}^{3} a_r(m_p) \exp\left(j \sum_{p=1}^{3} \phi(m_p)\right) \delta\left(\omega - \left(\sum_{p=1}^{3} i(m_p)\right)\omega_0\right) \qquad (5-13)$$

where

$$\mathbf{a}_r = |\, U_r(j\mathbf{i}\omega_0)\,| \qquad r = 1,2,3 \qquad \Phi = \angle U_1(j\mathbf{i}\omega_0) = \angle U_2(j\mathbf{i}\omega_0) = \angle U_3(j\mathbf{i}\omega_0)$$

and

$$S_{3,r}(j\omega_k) = Y_{3,r}(j\mathbf{i}\omega_0) \qquad \mathbf{i} = [\,1 \quad 3 \quad 5 \quad \ldots \quad 39\,] \qquad (5-14)$$

Since measurements were made at three amplitudes it was possible to make three combinations for use in the compensation technique. The FRF's estimated with uncompensated and compensated data for one amplitude pair are shown in Figure 5-12. The nonlinear disturbance adds a considerable bias to the magnitude and phase of the FRF's estimated with the uncompensated data, which increases with the signal amplitude. This bias is greatly reduced for the compensated FRF, which lies very close to the linear estimate.

The Fourier coefficients were then used as data for parametric estimation. The frequency-domain estimator was used to fit a first-order linear model to the uncompensated and compensated data. The results are given in Table 5-8, along with the complex error of the FRF.

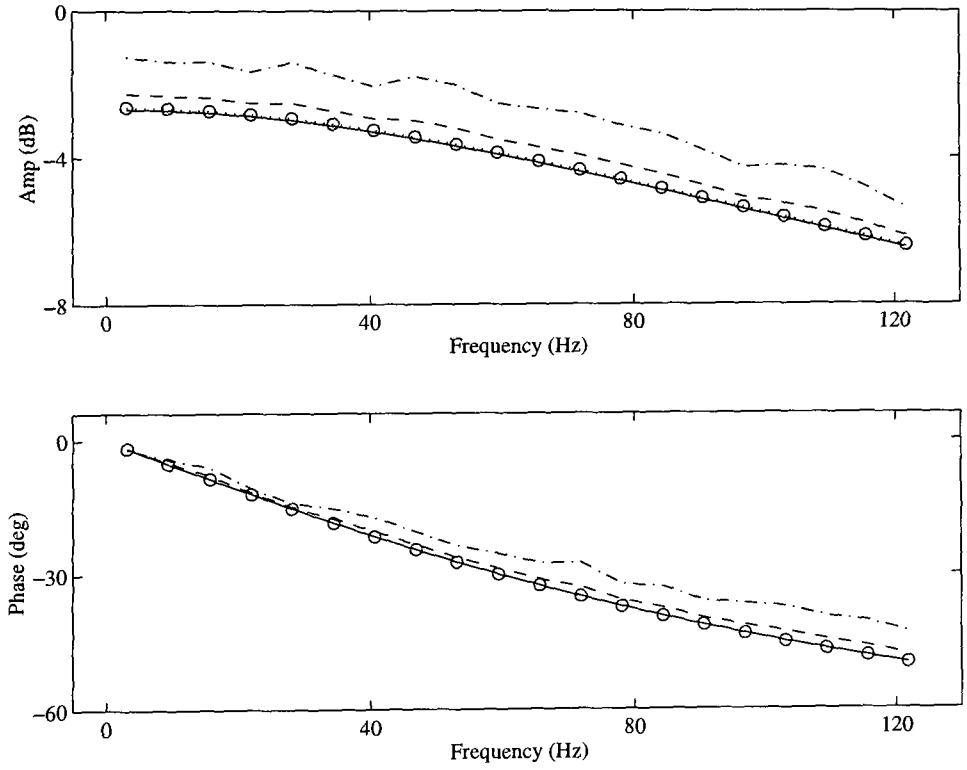

Figure 5-12. Estimates of FRF, with and without compensation. True FRF (solid), at amplitude 2 (dashed), at amplitude 3 (dash-dot) and compensated (dotted-circle).

It is seen that E increases with signal amplitude, as does the bias on the estimated poles. Using the compensated data greatly reduces the complex error and the bias on the pole is reduced to a very small value. The source of the remaining bias was also investigated, by using two of the data sets from the cubic compensation in a further compensation step for

TABLE 5-8

ERROR IN FRF, THE ESTIMATED POLE POSITIONS AND THEIR BIAS

Signal	E	Pole (s)	Bias
Amplitude 1	0.0057	-647.68	1.12
Amplitude 2	0.0122	-655.18	8.62
Amplitude 3	0.0340	-679.74	33.18
Compensated 1&2	0.0051	-646.94	0.38
Compensated 2&3	0.0053	-647.01	0.45
Compensated 1&3	0.0051	-646.93	0.37

a fifth-order term. The improvement in the FRF with this additional step was negligible, since the complex error remained at 0.0051, indicating that no significant fifth-order term was present.

Thus by testing at two amplitudes using an odd harmonic signal and applying a straightforward compensation technique it is possible to ensure that the linear estimate is unaffected by nonlinear contributions up to the fourth-order. Tests were performed at three amplitudes in this experiment, in order to allow the possible contribution of a fifth-order nonlinear term to be investigated.

In order to illustrate the technique, the comparison made in this section was between the "known" linear FRF and the compensated and uncompensated FRF's. In practice, the influence of the nonlinearity on the FRF estimate can be quantified by calculating the complex error E between the uncompensated and compensated frequency responses.

5.4 Analysis of the Residuals

Once a parametric linear model has been fitted to the frequency data, further information can be obtained from the frequency-domain residuals, calculated between the measured FRF and the model frequency response. Any modelling errors which are present stem from one of three sources: stochastic effects, under-modelling of the linear dynamics or unmodelled nonlinear dynamics.

Schoukens *et al.* (1998) have proposed a technique to distinguish between these errors, based on exciting the system with a random phased broad-band multisine and examining the normalised autocorrelation of the residuals. (The same paper was discussed in Chapter 4, in relation to the use of NID signals to measure the RLDS.) The normalised autocorrelation is defined as

$$R_{ee}(j\omega_m) = \frac{1}{(F-m)} \sum_{k=1}^{F-m} \frac{(\hat{H}(j\omega_k) - H(j\omega_k, \mathbf{p}))\,(\hat{H}(j\omega_{k+m}) - H(j\omega_{k+m}, \mathbf{p}))^*}{\sigma(j\omega_k)\,\sigma(j\omega_{k+m})} - \delta(m-0) \tag{5-15}$$

where $\hat{H}(j\omega_k)$ is the estimated FRF, $\sigma(j\omega_k)$ its standard deviation and $H(j\omega_k, \mathbf{p})$ the frequency response of the estimated model. The autocorrelation is a complex quantity, only the magnitude of which is of interest. Appropriate statistical bounds for the magnitude can be derived if R_{ee} is assumed to have a complex normal distribution. If there are no modelling errors then R_{ee} should be zero for all values of lag m. Taking into consideration the estimator cost function K, the residuals and R_{ee}, there are four possible scenarios:

Correct Model. The cost function is small, approaching $K_{\min}$, the residuals are small and R_{ee} is close to zero for all lags.

Stochastic Effects. The cost function is small, the residuals are large but R_{ee} is small, with a sharp form concentrated at $R_{ee}(0)$. The cost function is reduced by the large noise variances in the denominator of equation (2-37). The residuals are large due to the scatter of the FRF estimates but they are uncorrelated, resulting in small values of R_{ee} at all but the zero lag. The values of R_{ee} are further reduced by the large value of standard deviation in the numerator of equation (5-15).

Under-modelling. The cost function is large, the residuals are large and R_{ee} is also large, with significant values at non-zero lags. The broad shape of R_{ee} indicates that the residuals are correlated. This suggests that the difference between the model and the data is smoothly varying, which points to unmodelled linear dynamics.

Nonlinear Distortion. The cost function is large, the residuals are large and R_{ee} is also large, with a sharp form concentrated at $R_{ee}(0)$. The shape of R_{ee} indicates that the residuals are uncorrelated. This suggests that the model errors are due to nonlinear

distortions, so that increasing the order of the linear model will not significantly improve the fit. At this point it may be necessary to repeat the experiment and successively apply the compensation technique of Section 5.3, in order to remove the influence of the nonlinearities from the data.

Application of the residual analysis will be illustrated in Chapter 9, where it will be used to assess the influence of nonlinearities on the linear models of a gas turbine, estimated using consecutive and odd harmonic multisines.

5.5 An Identification Scheme

The techniques described in this chapter can be combined into one overall identification scheme, which takes account of nonlinear as well as stochastic errors in linear system identification. The suggested approach is shown in Figure 5-13, which aims to detect and quantify the degree of nonlinearity, remove its influence if necessary, and finally to iteratively estimate a linear model, while monitoring the cost function and residuals. If the estimated linear model is not satisfactory it may be necessary to perform, or repeat, the nonlinear compensation steps. If no improvement is obtained after repeated testing and compensation then a linear modelling approach is not appropriate to the problem.

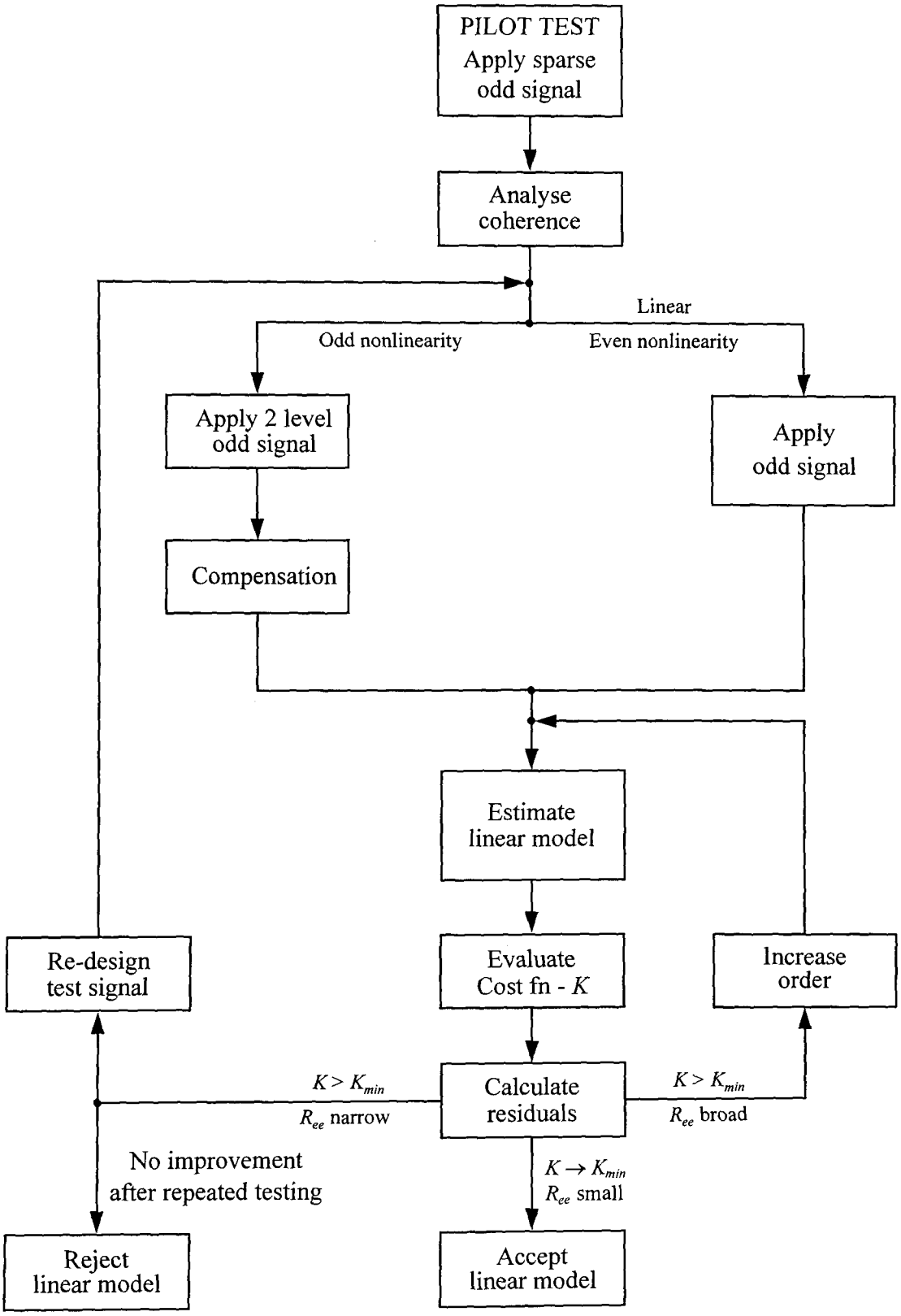

Figure 5-13. An identification scheme.

5.6 Conclusions

This chapter has examined the influence of various nonlinearities on a range of multisine signals. In contrast to Chapter 4, the aim was to isolate the linear component of the system response and not to obtain the best linear approximation of the overall nonlinear system. The linear component of the system response is independent of the input signal amplitude, while the linear approximation of the overall nonlinear system is strictly amplitude dependent.

A comparison was made between signals with varying numbers of harmonics but equal total power. The effect of the interaction of the Type I and Type II contributions defined in Chapter 4 was studied through simulation, for consecutive, odd, prime, odd-odd and NID harmonic multisines. If the phases of the input harmonics are set to zero, both the Type I and Type II contributions generated by the nonlinearity will be in-phase, which generates the maximum possible error. If random phases are used then the Type II contributions will also have randomly varying phases and their influence will be cancelled out as the number of experiments is increased. With low CF phases the Type II contributions tend to counteract the Type I contributions, resulting in a smaller overall nonlinear error. This effect is most significant with signals which generate many Type II contributions at the test frequencies.

For this reason, it is concluded that there is no benefit in using prime or NID harmonic signals if the aim is to *reduce* the nonlinear influence at the test frequencies to the minimum possible for a given input power. The work of Suki and Lutchen on designing NID-type signals for physiological measurements was discussed in this context. The key factor in reducing the influence of odd-order nonlinearities is the signal CF. An odd harmonic multisine is therefore recommended for general testing, since it is immune to even-order nonlinearities and amenable to CF minimisation. The simulation results were confirmed by practical tests, using controlled nonlinear disturbances, which examined the influence of second- and third-order nonlinearities, along with input and output saturation, on a range of multisines.

The presence of Type I contributions means that a nonlinear error is inherent with any odd-order nonlinearity and a compensation technique must be employed if its influence is to be completely eliminated. A method was proposed, based on testing the system at two amplitudes with the same multisine, and practical results were presented which illustrated its application. It was shown that the nonlinearity introduced an error into the estimated

FRF and a bias into the estimate poles, which were eliminated through application of the two-level technique. The advantage of this approach is that it can be applied without the need to specify a particular nonlinear model. The influence of the nonlinearity can be quantified by calculating the complex error between the uncompensated and compensated FRF's.

An overall identification scheme was then proposed for identifying linear models in the presence of nonlinear distortions. The steps involved include: an initial pilot test using a sparse odd signal, further testing using an odd harmonic signal, elimination of the nonlinear effects using a two level technique and, finally, the estimation of linear models. The estimator cost function and the frequency-domain residuals are used to quantify the nonlinear influence on the models. A systematic application of this approach will ensure the identification of high quality linear models, from which the nonlinear influence has been excluded, or a rejection of the linear modelling approach.

Chapter VI

Measuring Nonlinear Volterra Kernels

Abstract — *The frequency-domain measurement of the Volterra kernels of a nonlinear system using periodic multisine signals is now a practical possibility. An analysis is presented of the harmonic output of a Volterra kernel when excited with a multisine input, which lays the basis for the design of such signals. This is followed by a review of previous work in this area, after which a range of new periodic signals is defined. The minimisation of the signal crest factors is then examined, along with the practical problems associated with their application. Practical results are presented which illustrate the application of the signals to testing a known nonlinear circuit and a servo motor system.*

6.1 Introduction

There has been a steady growth in the study and modelling of nonlinear systems in recent years. Modern computer systems allow data to be gathered for longer periods, at faster speeds, and their enhanced computational power makes the analysis and identification of nonlinearities a practical possibility. In particular, it is now possible to move beyond the use of random inputs and correlation techniques for the measurement of Volterra or Wiener kernels. This chapter examines the specific problem of designing periodic test signals to directly measure the Volterra kernels of a nonlinear system.

The Volterra functional series representation of a nonlinear system was among the first to be postulated and has also been one of the most widely studied. It has found particular application in the analysis of communication systems (Narayanan (1970) and Bussgang *et al.* (1974)) and of mechanical structures (Gifford and Tomlinson, 1989). The related Wiener models have found particular application to biological systems (Hung and Stark (1977) and Marmarelis and Marmarelis (1978)). An excellent overview of the different techniques and applications involved has been presented by Billings (1980).

The Volterra structure has the attractive property that it can be seen as a natural extension of linear system theory. As such, the Volterra kernels have a direct physical significance and can often be given a physical interpretation, or be related to the system's constituent elements. The limitations of the Volterra structure are well documented, and these include the inability to model nonlinearities with discontinuous derivatives such hysteresis, dead-zone or backlash.

The chapter begins with a brief review of Volterra theory, followed by a discussion of the output frequency properties of a functional series nonlinearity when driven by a periodic multiharmonic input. There then follows a review of work previously conducted on identification of Volterra kernels, with particular emphasis on earlier approaches to designing signals of the type described in this chapter.

The results of the current study are then presented and a family of signals outlined. The importance of minimising the crest factor of the signals is discussed along with a number of important practical considerations when testing nonlinear systems. Finally, the signals are applied to a simulation example, followed by the testing of a known nonlinear circuit, made up of linear systems and multipliers, and a servo motor system.

6.2 Volterra Models

The Volterra functional series is a direct generalisation of the convolution integral description of linear systems (Eykhoff, 1974). For a causal, time-invariant system it can be expressed as

$$y(t) = \int_{-\infty}^{\infty} h_1(\tau)\, u(t-\tau)\, d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(\tau_1,\tau_2)\, u(t-\tau_1)\, u(t-\tau_2)\, d\tau_1 d\tau_2 + \ldots$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1,\tau_2,\ldots,\tau_n)\, u(t-\tau_1)\, u(t-\tau_2) \ldots u(t-\tau_n)\, d\tau_1 d\tau_2 \ldots d\tau_n \qquad (6-1)$$

where $h_1(\tau)$, $h_2(\tau_1,\tau_2)$ and $h_n(\tau_1,\tau_2, \ldots ,\tau_n)$ are known as the linear, second- and nth-order *kernels* respectively. The correspondence with the power series, dealt with previously in Chapter 4, is immediately apparent. If the process has no dynamics, or the input signal frequencies approach dc, then the functional series of equation (6-1) reduces to a power series in $u(t)$.

According to Fréchets theorem (Billings, 1980) an infinite Volterra series can represent the input-output relation of any time-invariant continuous nonlinear process. Boyd *et al.* (1984) have shown that if the process has noninfinite memory, termed *fading memory*, it can be represented by a finite Volterra series. In practice, a truncated series is usually employed, since the manipulation of higher-order kernels presents many difficulties.

The kernels suffer from a lack of uniqueness, in that several distinct nth-order kernels can give the same output for the same input. This problem can be overcome by symmetrising the kernels, defined as

$$h_n(\tau_1,\tau_2,\ldots,\tau_n)_{sym} \Leftrightarrow \frac{1}{n!}\{h_n(\tau_1,\tau_2,\ldots,\tau_n)_{asym} + \ldots + h_n(\tau_n,\tau_{n-1},\ldots,\tau_1)_{asym}\} \qquad (6-2)$$

where a symmetrical kernel is obtained by making all possible combinations of the arguments of the asymmetric kernel. The uniqueness of a symmetric kernel is guaranteed, as discussed by Boyd *et al.* (1984).

The transformation of the time-domain kernels into the s-domain is facilitated by the introduction of an associated equation for an nth-order functional, where the kernel is a function of new variables $t_1, t_2, \ldots, t_n$ instead of a single variable t.

$$y_{(n)}(t_1, t_2, \ldots, t_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \ldots, \tau_n)\, u(t_1 - \tau_1)\, u(t_2 - \tau_2) \ldots u(t_n - \tau_n)\, d\tau_1 d\tau_2 \ldots d\tau_n \tag{6-3}$$

The kernel can then be directly transformed by means of the multidimensional Laplace transform, which is defined as

$$H_n(s_1, s_2, \ldots, s_n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h_n(\tau_1, \tau_2, \ldots, \tau_n) \exp\left[-\sum_{i=1}^{n} s_i \tau_i\right] d\tau_1\, d\tau_2 \ldots d\tau_n \tag{6-4}$$

where $H_n(s)$ is termed the nth-order transfer function, or *s-domain kernel*, of the system. The multidimensional s-domain output of an nth-order kernel can be expressed as

$$Y_n(s_1, s_2, \ldots, s_n) = H_n(s_1, s_2, \ldots, s_n)\, U(s_1)\, U(s_2) \ldots U(s_n) \tag{6-5}$$

The elegance of the multidimensional Laplace transform is clearly illustrated, in that the complex multiple integrals of the time-domain expression have been replaced by straightforward multiplications. The symmetry properties of the s-domain kernels are analogous to those of the time domain and these will be exploited when measuring and plotting the kernels.

$$H_n(s_1, s_2, \ldots, s_n)_{sym} \Leftrightarrow \frac{1}{n!} \{H_n(s_1, s_2, \ldots, s_n)_{asym} + \ldots + H_n(s_n, s_{n-1}, \ldots, s_1)_{asym}\} \tag{6-6}$$

The Volterra representation lends itself naturally to the analysis of systems composed of linear elements and multipliers. For example, the transfer function of the system shown in Figure 6-1 can be expressed as a second-order kernel, where $\{\bullet\}_{sym}$ indicates a symmetrised transfer function.

$$H_2(s_1, s_2) = \{ A(s_1)\, B(s_2) \}_{sym} \tag{6-7}$$

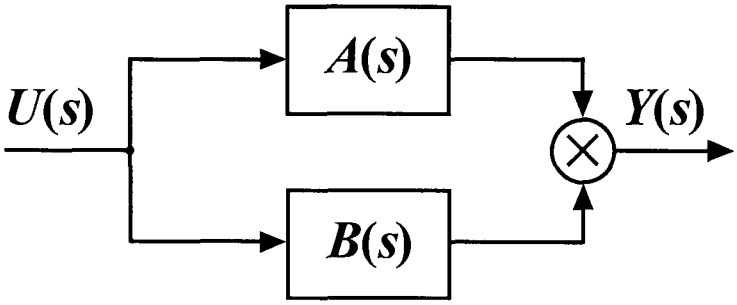

Figure 6-1. Nonlinear system composed of two linear systems and a multiplier.

The *higher order frequency response function* (HOFRF) of this kernel can be evaluated in a straightforward manner by utilising the symmetry properties of equation (6-6) and substituting $j\omega$ for s. If $A(s)$ is a first-order, low-pass system and $B(s)$ a resonant second-order system, with transfer functions

$$A(s)=\frac{1}{6s+1} \qquad B(s)=\frac{1}{3s^2+0.8s+1} \tag{6-8}$$

then the resulting Volterra kernel will have the two-dimensional amplitude plot shown in Figure 6-2.

This is expressed as a contour plot in Figure 6-3, where the symmetry along the $f_1 = f_2$ and $f_1 = -f_2$ diagonals is clearly visible. This symmetry means that the kernel is uniquely described by the shaded region in Figure 6-3. This frequency-domain representation of the kernels is greatly preferred, since it is far easier to interpret than the multidimensional impulse response of the equivalent time-domain kernel.

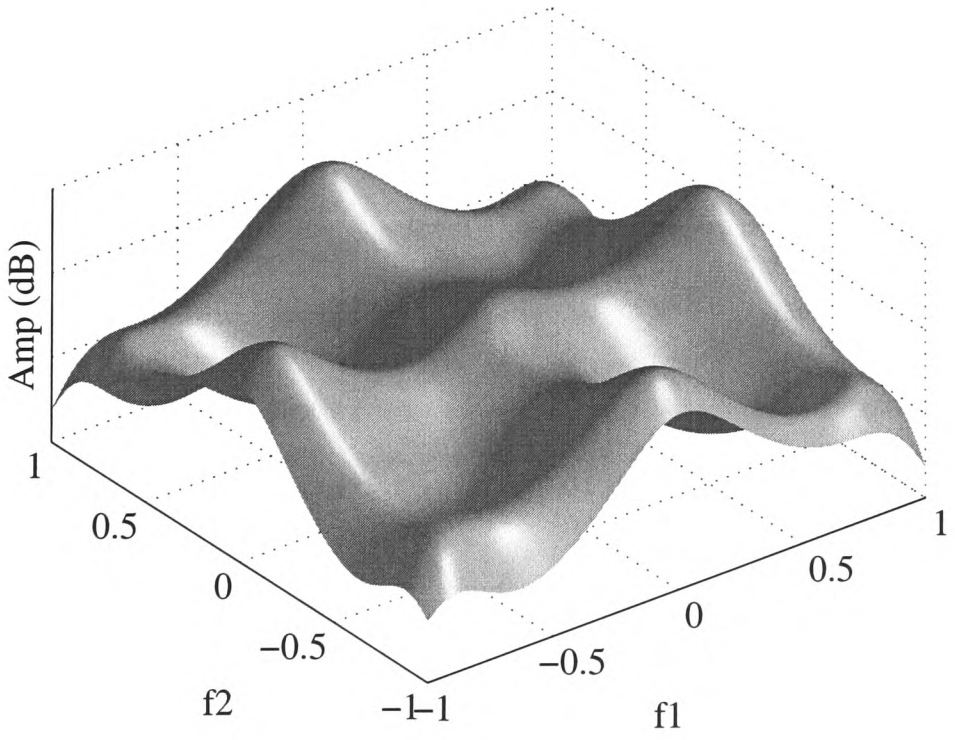

Figure 6-2. Amplitude of $H_2(s_1,s_2)$.

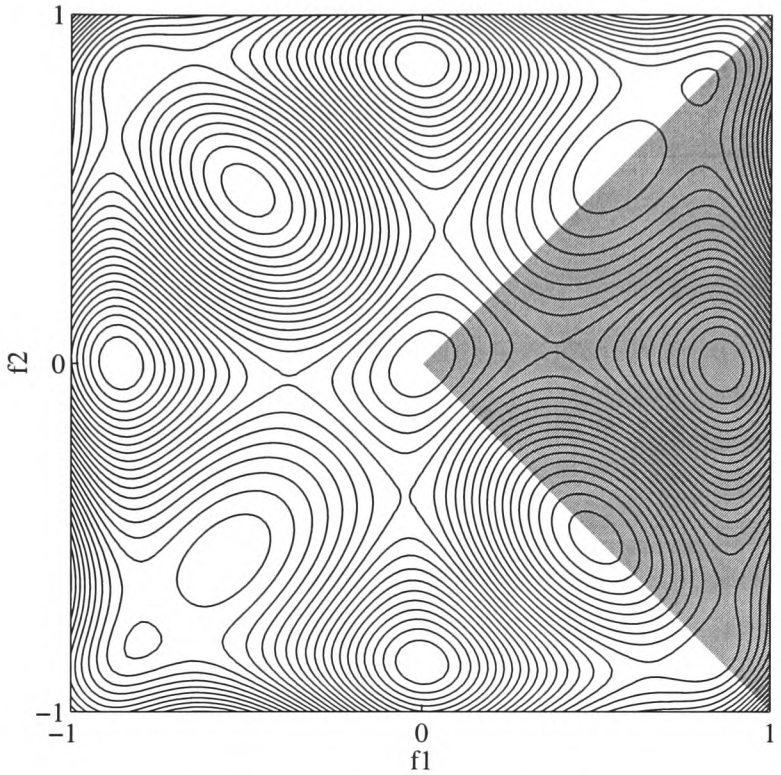

Figure 6-3. Contour plot of $| H_2(s_1,s_2) |$, showing unique region shaded.

6.3 Harmonic Analysis

The periodic steady-state theorem for Volterra models subject to multiharmonic inputs was derived by Boyd *et al.* (1984) and states that: "If the input u is periodic with period T for $t \geq 0$ then the output Nu approaches a steady state, also periodic with period T." The steady-state multidimensional frequency output of a nth-order kernel can thus be obtained directly from equation (6-5)

$$Y_{(n)}(j\omega_1, j\omega_2, \ldots, j\omega_n) = H_n(j\omega_1, j\omega_2, \ldots, j\omega_n)\, U(j\omega_1)\, U(j\omega_2) \ldots U(j\omega_n) \qquad (6-9)$$

In order to obtain the actual output spectrum $Y_n(j\omega)$ of the kernel, the multidimensional output of equation (6-9) must be collapsed into a single dimension. This transformation can be achieved by a technique known as the *association of variables*, which is described in detail by Schetzen (1980, pp. 104-108), and results in the following relation

$$Y_n(j\omega) = \frac{1}{(2\pi)^{n-1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Y_{(n)}(j\omega - j\mu_1, j\mu_1 - j\mu_2, \ldots, j\mu_{n-1})\, d\mu_1\, d\mu_2 \ldots d\mu_{n-1} \tag{6-10}$$

where

$$\mu_x = \omega_{x+1} + \omega_{x+2} \qquad x = 1, 2, \ldots, n-2$$

$$\mu_{n-1} = \omega_n$$

This equation is an (n-1)-dimensional convolution, which is generally difficult to evaluate. The convolution means that a single cosine applied to the kernel input will not propagate through the system independently of cosines at other frequencies. If the signal applied to a Volterra kernel of order n is a multisine of F cosines, with dc omitted and a double-sided spectrum

$$U(j\omega) = \sum_{k=-F}^{F} a(k)\, e^{j\phi(k)}\, \delta(\omega - i(k)\omega_0) \tag{6-11}$$

where

$$a(-k) = a(k)$$

$$i(-k) = -i(k) \qquad k = 1 \ldots F$$

$$\phi(-k) = -\phi(k)$$

then, following the analysis outlined in Chapter 4, the output can be expressed as the weighted discrete convolution of the spectrum, in effect the repeated convolution of impulse trains.

$$Y_n(j\omega) = \sum_{\substack{m_1=-F \\ m_1 \neq 0}}^{F} \sum_{\substack{m_2=-F \\ m_2 \neq 0}}^{F} \cdots \sum_{\substack{m_n=-F \\ m_n \neq 0}}^{F} H_n(j\omega_0[i(m_1), i(m_2), \ldots i(m_n)])$$

$$\times \left(\prod_{k=1}^{n} a(m_k) \right) \exp\left(j \sum_{k=1}^{n} \phi(m_k) \right) \delta\left(\omega - \left(\sum_{k=1}^{n} i(m_k) \right) \omega_0 \right) \tag{6-12}$$

It is clear from equation (6-12) that the harmonics generated at the output will consist of all the permutations of sums of n input harmonics. It is quite possible, and indeed likely,

that sums of different input harmonics will produce output contributions at the same frequency. Following Billings and Tsang (1989), an n-th order module vector $\mathbf{m}$ can be defined for the $2F$ input frequencies

$$\mathbf{m} = [\, m_{-F}\ m_{-F+1}\ \ldots\ m_{F-1}\ m_F \,] \tag{6-13}$$

where m_i are non-negative integers ($m_i \geq 0$) and

$$\sum_{-F}^{F} m_i = n \tag{6-14}$$

The number of module vectors which obey the constraint of equation (6-14) is given by the number of combinations of the $2F$ input frequencies, taken n at a time, with repetition

$$q = \frac{(2F+n-1)!}{n!\,(2F-1)!} \tag{6-15}$$

which allows a $q \times 2F$ module matrix $\mathbf{M}$ to be defined

$$\sum_{r=1}^{q} \sum_{c=1}^{2F} \mathbf{M}(r,c) = q \times n \tag{6-16}$$

The output harmonic vector will then be

$$\mathbf{o} = \mathbf{M}\,\mathbf{i}^T \tag{6-17}$$

For example, consider an input composed of the first and second harmonics applied to a second-order kernel, then

$$\mathbf{m} = [\, m_{-2}\ m_{-1}\ m_1\ m_2 \,] \qquad \sum_{-F}^{F} m_i = 2 \qquad q = \frac{5!}{2!\,3!} = 10 \tag{6-18}$$

$$\mathbf{i} = [\, -2\ -1\ 1\ 2 \,] \tag{6-19}$$

Hence

$$\mathbf{o} = \mathbf{M}\mathbf{i}^T = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -1 \\ 1 \\ 2 \end{bmatrix} \tag{6-20}$$

$$\mathbf{o}^T = [-4\ -3\ -2\ -1\ 0\ 0\ 1\ 2\ 3\ 4] \tag{6-21}$$

It can be seen that contributions will be generated at dc and the first, second, third and fourth harmonics. The contributions at the first and second harmonics will coincide with the output of the linear kernel but the contributions at the third and fourth harmonics are distinct. From (6-12), their complex amplitudes are directly related to the two input harmonics which gave rise to them and to points on the HOFRF

$$\begin{aligned} Y_2(j3\omega_0) &= H_2(j\omega_0, j2\omega_0)\, U(j\omega_0)\, U(j2\omega_0) + H_2(j2\omega_0, j\omega_0)\, U(j2\omega_0)\, U(j\omega_0) \\ Y_2(j4\omega_0) &= H_2(j2\omega_0, j2\omega_0)\, U(j2\omega_0)\, U(j2\omega_0) \end{aligned} \tag{6-22}$$

If the arguments of $H_2(j\omega_1, j\omega_2)$ are different, they can be interchanged without affecting the resulting output frequency and hence two points on the HOFRF contribute to the output at that frequency. Thus the measurements of $Y(j3\omega_0)$ and $Y(j4\omega_0)$ allow three points on the HOFRF to be directly measured. The measured points are shown as crosses on the contour plot in Figure 6-4. It can be seen that output frequencies resulting from sums of the same positive input frequency allow measurement of points along the $f_1 = f_2$ diagonal. Sums of different frequencies each allow the measurement of two points on the HOFRF, which have the same amplitude and phase, due to the symmetry across the $f_1 = f_2$ diagonal.

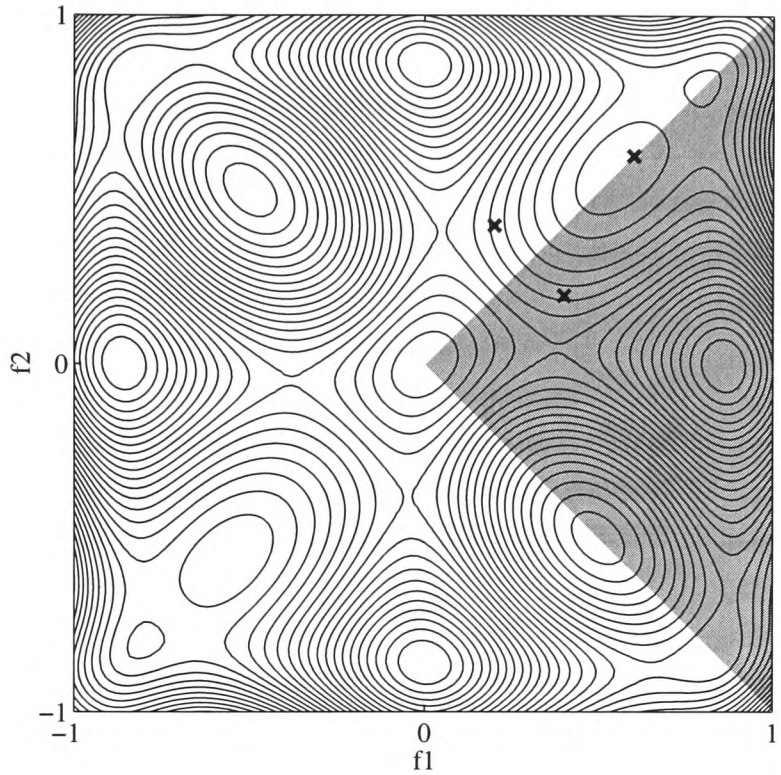

Figure 6-4. Contour plot of $|H_2(s_1,s_2)|$, showing the points measured by a two-tone signal (crosses).

Generalising this approach, it is possible to directly estimate points on the HOFRF of an nth-order kernel using those output contributions which result from a distinct sum of the input harmonics, and only those harmonics

$$\hat{H}_n(j\omega_1, j\omega_2, \ldots, j\omega_n) = \frac{Y\left(\sum_{i=1}^{n} j\omega_i\right)}{k \prod_{i=1}^{n} U(j\omega_i)} \qquad (6-23)$$

where the weighting k depends on the number of points on the kernel which correspond to different combinations of the same input harmonics.

For the specific example of the second-order kernel

$$\hat{H}_2(j\omega_1, j\omega_2) = \frac{Y(j\omega_1 + j\omega_2)}{k\ U(j\omega_1)\ U(j\omega_2)} \qquad (6-24)$$

where

$k = 1$ if $\omega_1 = \omega_2$

$k = 2$ if $\omega_1 \neq \omega_2$

Measurements of unique kernel points can be made in the shaded area of Figure 6-4, which is bounded by, but does not include, the $f_1 = -f_2$ diagonal. This is because the frequency combinations along this diagonal will all result in contributions at dc, which cannot be separately measured.

Since a large number of points are needed to plot a complete kernel, the use of two-tone signals is not very practical, because a large number of tests would have to be performed. For this reason signals are required which possess the same properties but contain a larger number of harmonics. The analysis presented in Chapter 4 showed that the output contributions generated by a nonlinear element can be divided into two basic types. Type I, or *harmonic* contributions, will always be generated whatever the harmonic content of the signal, with their number depending on the order of the kernel and the number of input harmonics.

For even-order kernels, Type I contributions result from the pairing of equal negative and positive harmonics, which result in contributions at dc. One pair of negative and positive harmonics for a second-order kernels, two pairs for a fourth-order kernel, and so on. For odd-order kernels Type I contributions result from the summation of pairs of equal negative and positive harmonics with one other input harmonic, resulting in output contributions at that input harmonic. Thus these Type I contributions present a problem in kernel measurement, which was mentioned above and will be discussed in greater detail in Section 6.5.1.

The remaining contributions, which were termed Type II or *interharmonic*, can be influenced by the selection of the input harmonics. It was shown in Chapter 4 that it is possible to design harmonic sets which do not generate any Type II contributions at the input harmonics, for a given order of nonlinearity. This implies that all the Type II

contributions must fall at excluded harmonics and that they can be directly measured. However, it was found that in many cases a number of different combinations of input harmonics would generate a contribution at the same output harmonic, which prevented the measurement of a distinct point on the Volterra kernel.

The design criterion must therefore be extended for the purpose of Volterra measurement, to include the stipulation that each Type II contribution must fall at a *distinct* output harmonic. The aim of this work can thus be summarised as the design of periodic multisine signals which allow the measurement of the maximum possible points on a Volterra kernel, for a given kernel order and number of input harmonics. The measurements can then be used to plot the nonparametric HOFRF or as data for the parametric estimation of s-domain kernels, as discussed by Lawrence (1981) and Schoukens *et al.* (1988).

6.4 Previous Work

Traditional methods of measuring Volterra kernels, or the related Wiener functionals, rely on the use of Gaussian input signals and correlation techniques (Lee and Schetzen (1965) and Schetzen (1980)). These approaches all share the uncertainty associated with the use of random signals and the introduction of systematic leakage errors when transforming the signals using the FFT. It is also difficult to ensure that the statistical properties of the finite length noise sequences used in practice are truly Gaussian. Very long test times are required in order to achieve accurate results, with anything between 10,000 and 100,000 points being measured (Korenberg, 1991). Consequently, the work conducted has been largely restricted to measuring first- and second-order kernels.

An alternative approach, using pseudo-random m-level sequences, has been studied by a number of authors (Ream, 1970). The use of deterministic signals allowed the test times to be greatly reduced but also caused bias problems in the correlation analysis. This problem has been extensively studied by Barker and various co-workers (see Barker and Davy, 1978), who eliminated the bias by selection of the most suitable sequences and applying a compensation method. Due to the difficulties involved, the application of m-level signals has been largely restricted to ternary sequences for the identification of second-order kernels.

All of the work described thus far has relied on certain properties of the input signals in order to obtain kernel estimates. This is also the chosen approach in this thesis, using periodic multisine signals with certain harmonics omitted. However, before proceeding to discuss such signals it is worthwhile to digress slightly and mention a completely different approach, recently developed by Billings and his co-workers. This method relies on estimating a nonlinear auto-regressive moving average (NARMAX) model and then using a frequency probing technique to arrive at a HOFRF, expressed in terms of gains and complex exponentials (Billings and Tsang (1989) and Zhang and Billings (1993)).

This approach has the great advantage that the only restriction which needs to be placed on the input signal, in order to estimate a NARMAX model, is that it sufficiently excites all the modes of the system. However, the quality of the resulting HOFRF will depend entirely on the quality of the discrete model on which it is based. This will in turn depend on the sampling frequency used and the experimental conditions. Hence, the method described in this chapter can be seen as complementary to the NARMAX approach, in that it allows the direct measurement of points on the HOFRF, which can then be used to validate the HOFRF obtained from the NARMAX models. This has obvious parallels with the use of a linear FRF to validate continuous and discrete linear models.

Returning to the design of input signals, the advances in microprocessor technology achieved by the late 1970's made it possible to tackle the measurement problem more directly, by applying specialised sum-of-sinusoid signals. Victor and Knight (1979) first proposed the use of signals composed of sums of incommensurate frequencies, which are frequencies which cannot be expressed in terms of a fundamental and integer multiples. They showed that such signals generate kernel output spectra in which each of the possible combinations of input frequencies fall at distinct output frequencies. A related paper by Victor (1979) demonstrated the improvement obtained with such signals, when compared to Gaussian noise, for the measurement of the first- and second-order kernels of the neural pathway of a cat retina.

The stated advantages of these signals were the improvement obtained in the signal to noise ratio and their deterministic nature; the main disadvantage being their lack of periodicity. Further work by Victor and Shapley (1980) addressed this problem by defining periodic signals composed of commensurate frequencies, made up of selected higher multiples of a fundamental frequency. These periodic signals were similarly applied to the measurement of the first- and second-order kernels of a cat retina.

The signal harmonics were selected either empirically, or according to simple equations, which result in a signal of ten frequencies with the harmonic vector shown in equation (6-25).

$$\mathbf{i} = [7 \quad 15 \quad 31 \quad 63 \quad 127 \quad 255 \quad 511 \quad 1023 \quad 2047 \quad 4095] \qquad (6-25)$$

The harmonic value is effectively doubling with each additional frequency, which means that the gaps between the harmonics become very large as the total number is increased. This uneven spacing leads to a poor coverage of the kernel at higher frequencies, as shown in Figure 6-5 by a plot of the points measured with this signal.

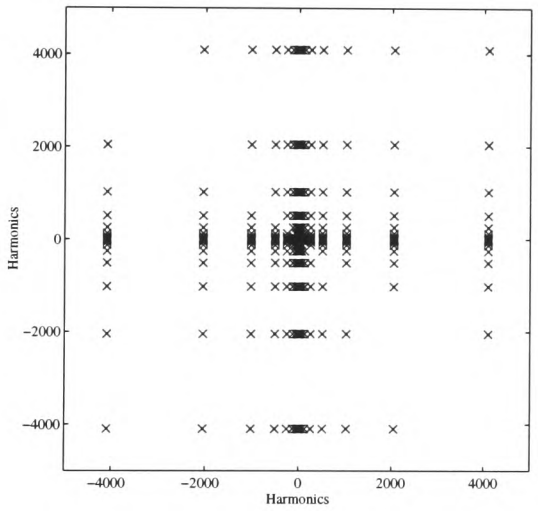

Figure 6-5. Kernel coverage plot for harmonic vector (6-25).

The next reported application of such signals was by Lawrence (1981), who designed periodic signals which possessed autocorrelation properties analogous to Gaussian noise, for a given order of nonlinearity. If white autocorrelation properties are required up to the third-order, the design procedure generates a ten frequency signal with the following harmonics

$$\mathbf{i} = [1 \quad 4 \quad 10 \quad 17 \quad 29 \quad 52 \quad 67 \quad 89 \quad 132 \quad 164] \qquad (6-26)$$

Lawrence showed that this signal can be used to measure points on the first- and second-order kernels, since no nonlinear contributions will fall at the input frequencies, and each of the Type II contributions generated by the second-order kernel will fall at a distinct output frequency. The signal spectrum is considerably more compact than the

signal designed using the Victor and Shapley approach, with the tenth input frequency occurring at the 164th harmonic. The points on the second-order kernel measured by this signal are shown in Figure 6-6.

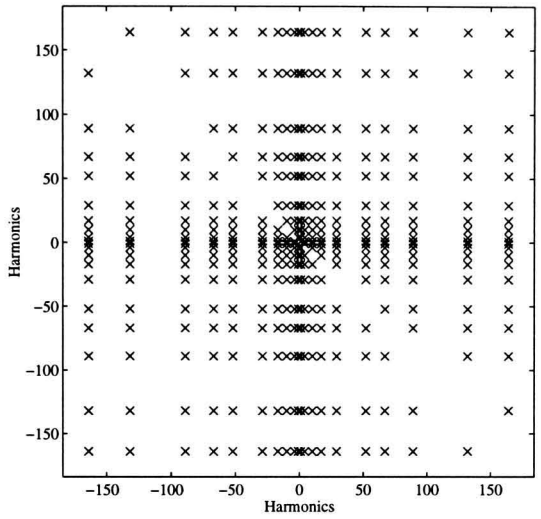

Figure 6-6. Kernel coverage plot for harmonic vector (6-26).

A signal with Gaussian autocorrelation properties up to the fifth order is required in order to measure distinct points on a third-order kernel. The harmonic vector of the signal presented by Lawrence is shown in equation (6-27) and is seen to have a very wide bandwidth, with large gaps between the higher signal harmonics.

$$\mathbf{i} = [8 \quad 17 \quad 31 \quad 66 \quad 143 \quad 273 \quad 508 \quad 1018 \quad 2046 \quad 4099] \qquad (6-27)$$

The problem was also addressed by Boyd *et al.* (1983), apparently independently of both Victor *et al.* and Lawrence, with the aim of measuring a second-order kernel. They postulated an explicit formula to construct signals made up of two harmonic sets, whereby the combinations of frequencies from the first and second sets were completely separate at the output. This approach leads to harmonic vectors such as that shown in equation (6-28).

$$\mathbf{i} = [3 \quad 6 \quad 9 \quad 10 \quad 12 \quad 15 \quad 18 \quad 21 \quad 24 \quad 27] \qquad (6-28)$$

While the signal is far more compact than harmonic vectors (6-25) and (6-26), it is clear that many contributions will be generated at the input frequencies. In addition, the design method only ensures that a certain number of the second-order contributions fall at

distinct output frequencies, which means that a large number of possible combinations are lost. The kernel coverage of this signal is shown in Figure 6-7, which shows the number of measurement points to be significantly less than with the Lawrence signal, while the spacing of the measurements is more regular. Such a signal was used by Boyd *et al.* to measure the second-order kernel of an electro-acoustic transducer.

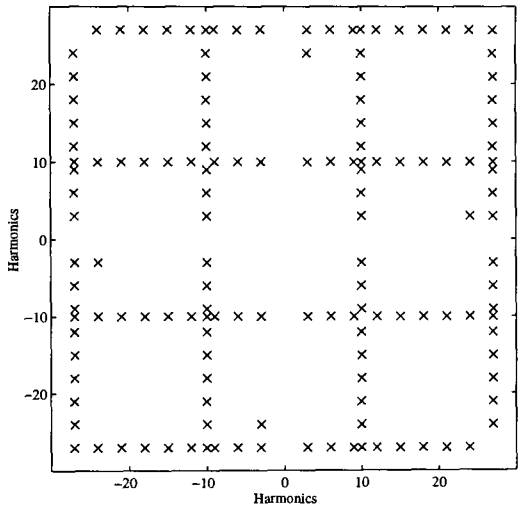

Figure 6-7. Kernel coverage plot for harmonic vector (6-28).

This approach was extended by Chua and Liao (1989) to the design of signals to measure higher-order kernels and also to estimating the highest significant kernel order (Chua and Liao, 1991). For the third-order case they postulated a signal made up of three sets of harmonics and found that an explicit formula for selecting these sets gave poor results. They then used a computer search to establish a series of rules for signal design, which give rise to harmonic vectors such as that shown in equation (6-29). These signals have the advantage of being very compact, with quite evenly spaced harmonics, and the disadvantage that many combinations of the input harmonics do not fall at distinct output frequencies and are thus lost.

$$\mathbf{i} = [80 \quad 81 \quad 90 \quad 160 \quad 162 \quad 180 \quad 240 \quad 243 \quad 270 \quad 320 \quad 324 \quad 360] \qquad (6-29)$$

Since the signals are made up of three harmonic sets of equal length, it is not possible to design a signal of 10 harmonics. The last ten harmonics of equation (6-29) will thus be used in later comparisons.

6.5 New Design Method

The aim of the current work was to design signals which overcame the drawbacks associated with each of the methods previously described. Namely, it was to generate signals which allowed the maximum possible points on a kernel to be measured in one test, while maintaining a near-even harmonic spacing and minimising the highest harmonic. Attempts at finding an analytical or numerical solution to this problem encountered the same difficulties as those experienced by Chua *et al.* and resulted in the use of an exhaustive search method.

A search algorithm was used to progressively build up harmonic vectors which generated no more than one Type II contribution at any excluded frequency and none at any of the input frequencies. The search was initially started at the fundamental, with each subsequent harmonic being tested for possible inclusion. This produced signals similar to those designed by Lawrence, which have the undesirable property that the gaps between the harmonics get progressively larger as the number of harmonics is increased.

It was found that the harmonic selection can be influenced by three design variables:

(1) The *start* harmonic for the search can be varied;

(2) Each time a suitable harmonic is found and added to the harmonic vector **i**, a *jump* of a certain number of harmonics can be imposed before re-commencing the search;

(3) The set of harmonics can be restricted to odd harmonics only.

The first two variables affect the spacing of the resulting vector, while the third ensures that the even and odd kernel contributions will only fall at even and odd harmonics respectively. The quality of the harmonic vectors generated by altering these variables was assessed by examining the maximum harmonic of each set and the spacing of the harmonics. Defining a vector of harmonic differences

$$\Delta\mathbf{i} = [\ \mathbf{i}(2) - \mathbf{i}(1) \quad \mathbf{i}(3) - \mathbf{i}(2) \quad \ldots \quad \mathbf{i}(F) - \mathbf{i}(F-1)\ ] \tag{6-30}$$

then the standard deviation of $\Delta\mathbf{i}$, which will be denoted as $\sigma_{\Delta i}$, can be used as a measure of the near-even spacing of **i**. An exhaustive computer search was carried out to find the

harmonic sets with the lowest max(**i**) and $\sigma_{\Delta i}$, by varying the start and jump values over a wide range. It was found that varying the start and jump values only produced a small improvement in max(**i**) but considerably influenced $\sigma_{\Delta i}$.

These observations led to a simplified design procedure based on first conducting an exhaustive search, in order to establish an approximate value for the maximum harmonic, and then repeating the search with larger jumps, in order to improve the harmonic spacing. The procedure can be summarised as follows:

(1) Select the number of input frequencies F.

(2) Search for a harmonic set $\mathbf{i}_1$, using initial values of *start* = 1 and *jump* = 1.

(3) Define $k_{max} = \max(\mathbf{i}_1) / F$. Select *jump* values close to [0.8 1 1.2] k_{max}.

(4) Select a range of *start* values, beginning either at a low harmonic or at a harmonic close to but less than k_{max}.

(5) Repeat the search with the new range of *start* and *jump* values.

(6) Select the signal with the minimum value of $K = \max(\mathbf{i}) + w\sigma_{\Delta i}$, where the weighting w is chosen to reflect the importance of obtaining a near-even harmonic vector.

Two options are given for the choice of start values in step (4). Using a low start value will mean that the first measurements are made close to the signal fundamental, which has the advantage that the fundamental will not have to be pushed down too far in order to measure at a given minimum frequency. However, using higher start values was found to generate vectors with better harmonic spacing. These points will be illustrated by the signals described in the next section. The design procedure was used to generate signals of ten harmonics, to be used for second- and third-order kernel measurements.

6.5.1 Second-order

The signals were initially built up from both even and odd harmonics, using low start values, and the best harmonic vector is given in equation (6-31).

$$\mathbf{i} = [1 \quad 8 \quad 18 \quad 29 \quad 41 \quad 56 \quad 80 \quad 102 \quad 133 \quad 146]$$

$$\text{Start} = 1,\ \text{Jump} = 7 \qquad \sigma_{\Delta i} = 7.8 \qquad (6-31)$$

Of the signals with higher start values, the following had the best harmonic set.

$$\mathbf{i} = [13 \quad 22 \quad 32 \quad 43 \quad 55 \quad 70 \quad 84 \quad 101 \quad 121 \quad 137]$$

$$\text{Start} = 13, \ \text{Jump} = 9 \qquad \sigma_{\Delta i} = 3.6 \qquad (6-32)$$

The disadvantage of this signal is that the first input frequency is at the 13th harmonic, which means that no measurements will be made below this point. The points on the kernel measured by harmonic vector (6-32) are shown in Figure 6-8, which shows a more even spread of measurement points than the Victor or Lawrence signals and an improved measurement density compared to the Boyd signal.

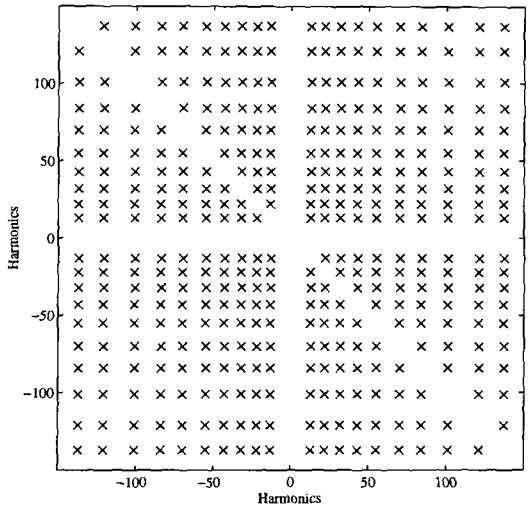

Figure 6-8. Kernel coverage plot for harmonic vector (6-32).

The problem with the above signals is that the points at which the second-order kernel is measured will also be corrupted by higher-order nonlinear contributions. If a weak nonlinearity is being measured, it may be possible to assume that contributions from the fourth and higher order kernels are negligible, but some account must be taken of the third-order contributions. Two options are available in this case, the first being to measure at two or more input signal amplitudes and use an interpolation technique to isolate the second-order contributions, which will be discussed in Section 6.5.3. Alternatively, the input signal harmonics can simply be restricted to an odd set. Signals of 10 harmonics were constructed using this approach, the harmonic vector of the most suitable is shown in equation (6-33).

$$\mathbf{i} = [3 \quad 13 \quad 25 \quad 43 \quad 57 \quad 77 \quad 119 \quad 155 \quad 203 \quad 227]$$

$$\text{Start} = 3, \ \text{Jump} = 10 \qquad \sigma_{\Delta i} = 13.8 \qquad (6-33)$$

In comparison with harmonic vector (6-32) the highest harmonic has increased from 137 to 227. This will mean a considerable increase in the overall test time, if the maximum harmonic is to be fixed at a given frequency. But the increase will be less than repeating the test two or more times with harmonic vector (6-32), particularly if the system settling time is taken into account. The harmonic spacing can be improved by using a higher start value, at the cost of a higher maximum value in this case.

$$\mathbf{i} = [27 \quad 43 \quad 61 \quad 83 \quad 103 \quad 127 \quad 155 \quad 185 \quad 217 \quad 263]$$

$$\text{Start} = 27, \ \text{Jump} = 16 \qquad \sigma_{\Delta i} = 9.2 \qquad (6-34)$$

Problem of Type I Contributions. This is illustrated in the principal diagonal of Figure 6-8. The gaps in the measurements are points corresponding to $f_1 = -f_2$ contributions, which all fall at dc and therefore cannot be separately measured. The only way to directly measure points along this diagonal is to use single sine inputs and measure the dc value at the output, which is a difficult and cumbersome process. If a higher measurement density is required around this diagonal a signal can be designed especially for this purpose, incorporating closely spaced harmonic pairs, in order to measure points close to the diagonal.

Such signals can be built by amending the design procedure, so that the jump imposed after a new harmonic is added to the set is alternated between a small and a large value. A signal of 20 harmonics is required if ten pairs of points are to be measured close to the diagonal. The harmonic vector of a suitable signal is given in equation (6-35), its spectrum is shown in Figure 6-9 and its kernel coverage in Figure 6-10.

$$\mathbf{i} = [3 \quad 5 \quad 75 \quad 76 \quad 158 \quad 162 \quad 246 \quad 253 \quad 338 \quad 347 \quad 436 \quad 447 \quad \ldots$$

$$540 \quad 552 \quad 642 \quad 655 \quad 772 \quad 787 \quad 907 \quad 921]$$

$$\text{Start} = 3, \ \text{Jump} = 1 \text{ and } 70 \qquad \sigma_{\Delta i} = 44.3\,. \qquad (6-35)$$

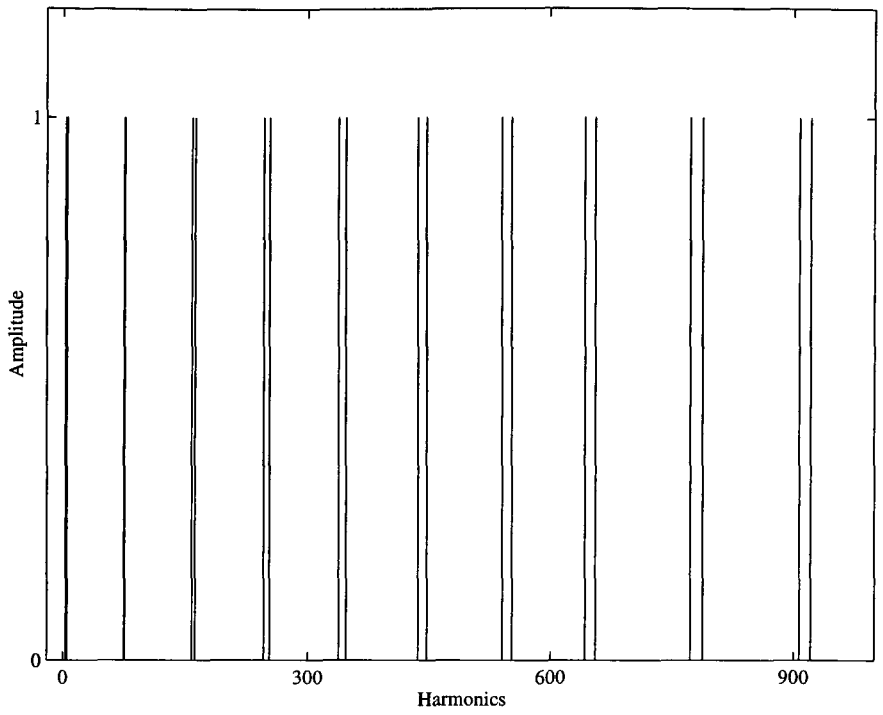

Figure 6-9. Amplitude spectrum of signal with harmonic vector (6-35).

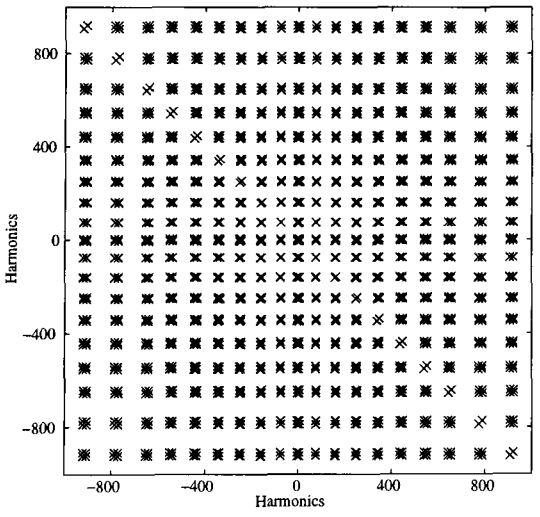

Figure 6-10. Kernel coverage plot for harmonic vector (6-35).

6.5.2 Third-order

When searching for harmonic sets suitable for measuring third-order kernels it is important not only to ensure that all the Type II contributions fall at distinct frequencies but also that none falls at dc. One of the drawbacks of the Chua *et al.* signals is that they generate a large number of contributions at dc. Since the different contributions which fall at dc cannot be separated, these measurement points are lost.

The best results for signals designed to measure third-order kernels will now be presented, using both low and high start values. For a consecutive harmonic set, the following signals were selected.

$$\mathbf{i} = [5 \quad 125 \quad 246 \quad 368 \quad 494 \quad 631 \quad 769 \quad 908 \quad 1067 \quad 1213]$$
$$\text{Start} = 5, \ \text{Jump} = 120 \qquad \sigma_{\Delta i} = 13.2 \qquad (6-36)$$

$$\mathbf{i} = [110 \quad 221 \quad 333 \quad 446 \quad 561 \quad 679 \quad 798 \quad 926 \quad 1056 \quad 1187]$$
$$\text{Start} = 110, \ \text{Jump} = 110 \qquad \sigma_{\Delta i} = 8.0 \qquad (6-37)$$

The problem with such signals is that some points at which the third-order kernel is measured will also be points at which second-order contributions fall. This problem can be overcome by once again restricting the search harmonics to an odd set. This results in the harmonic vectors shown in equations (6-38) and (6-39), where the highest harmonics have effectively doubled, when compared to vectors (6-36) and (6-37). It can be seen that the best overall results are obtained using signals with high start values.

$$\mathbf{i} = [41 \quad 291 \quad 543 \quad 797 \quad 1055 \quad 1319 \quad 1585 \quad 1927 \quad 2229 \quad 2543]$$
$$\text{Start} = 41, \ \text{Jump} = 250 \qquad \sigma_{\Delta i} = 33.1 \qquad (6-38)$$

$$\mathbf{i} = [241 \quad 451 \quad 663 \quad 877 \quad 1095 \quad 1319 \quad 1581 \quad 1817 \quad 2109 \quad 2347]$$
$$\text{Start} = 241, \ \text{Jump} = 210 \qquad \sigma_{\Delta i} = 27.4 \qquad (6-39)$$

Plotting third-order kernels is a somewhat involved process, since the amplitude and phase are now a function of three frequencies. The method most commonly employed is to successively fix each frequency and show a series of three-dimensional mesh plots.

Since each plot then shows only certain aspects of the overall kernel topography, this makes interpretation very difficult. This problem can be overcome by plotting the kernel as a cube, with the amplitude or phase shown as variations in colour, as illustrated by Weiss *et al.* (1996).

The Type I contributions cause the same problem in the measurement of third-order kernels, in that contributions from points in $H_3(f_1,f_2,f_3)$ where $f_1 = -f_2$, or $f_1 = -f_3$, or $f_2 = -f_3$, will fall at the input frequencies and cannot be separately measured. This problem is illustrated in Figure 6-11, which shows the unmeasured points in the three-dimensional kernel. Three perpendicular planes are shown, which correspond to kernel values when one frequency is at dc. There are also three diagonal planes corresponding to the unmeasured Type I contributions. This can be understood by considering each face of the cube as a surface where one of the three frequencies is constant. A diagonal across this surface then corresponds to the $f_1 = -f_2$ points of the two varying frequencies. Each of these diagonals extends through the cube, creating three planes of unmeasured points.

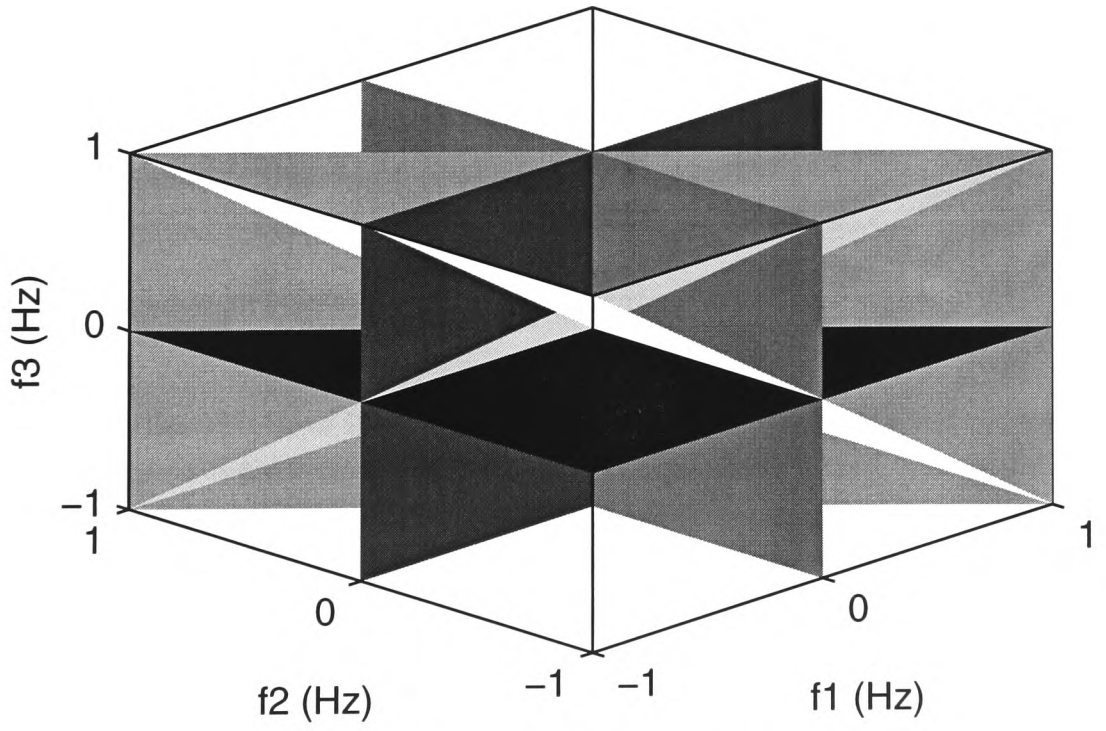

Figure 6-11. Unmeasured points in a third-order kernel, (from Weiss *et al.* (1996)).

6.5.3 Higher Order Kernels

Signals for measuring kernels of higher order can be designed by extension of the technique described above. However, it will no longer be possible to separate the kernel contributions completely, since all nth-order output frequencies are present in the $(n+2)$th-order output, as shown by Chua and Ng (1979). The separation of the nth- and $(n+2)$th-order contributions can be achieved by repeating the test at a number of input amplitudes and employing an interpolation technique, such as constructing a Vandermonde matrix and solving for the coefficient of each power term, as described in detail by Boyd *et al.* (1983) and Chua and Liao (1989).

6.5.4 Comparison of Signals

The new signals can be compared to the previous designs in terms of the number of unique kernel points which are measured, the maximum harmonic of each signal and the harmonic spacing. The maximum number of unique points which can be measured using a signal of F cosines can be found using equation (6-15), by calculating the total number of contributions, subtracting the number of Type I contributions and then dividing by two to exclude the negative frequencies. For second- and third-order kernels this leads to

$$\text{Maximum 2nd-order points} = \left(\frac{(2F+1)\,2F}{2!} - F \right) / 2 = F^2 \qquad (6-40)$$

$$\text{Maximum 3rd-order points} = \left(\frac{(2F+2)\,(2F+1)\,2F}{3!} - 2F^2 \right) / 2 = \frac{2F^3 + F}{3} \qquad (6-41)$$

This means that a signal of 10 cosines will enable the measurement of a maximum of 100 points on a second-order kernel and 670 points on a third-order kernel. A summary of the signals designed to measure second-order kernels is presented in Table 6-1. It is immediately apparent that the sparsity of the Victor and Shapley signal is far in excess of that required. Indeed, the signal has a higher maximum harmonic than most of the signals designed to measure points on a third-order kernel.

The Boyd *et al.* signal is the most compact but it only measures a small number of points on the second-order kernel, while the Lawrence and Evans signals each measure the maximum 100. The Evans signal has the advantage that it is more evenly spaced, with a lower maximum harmonic. However, nearly all the second-order measurement points of

these three signals are corrupted by third-order contributions. This effect is eliminated with the Evans odd signal, which simultaneously measures 100 uncorrupted points on the second-order kernel and 70 on the third.

TABLE 6-1

SUMMARY OF SECOND-ORDER SIGNAL DESIGNS

Eqn.	Signal	Harmonics			Kernel Measurement Points			
		Min	Max	$\sigma_{\Delta i}$	2nd-Order	Corrupt by 3rd	3rd-Order	Corrupt by 2nd
6-25	Victor *et al.*	7	4095	683.3	100	0	405	0
6-26	Lawrence 2nd	1	164	13.2	100	94	100	13
6-28	Boyd *et al.*	3	27	0.7	21	21	5	0
6-32	Evans 2nd	13	137	3.6	100	95	89	14
6-33	Evans 2nd odd	3	227	13.8	100	0	70	0

Table 6-2 summarises the signal designs for measuring third-order kernels. The Chua signal has the advantage of compactness but only measures 234 points on the third-order kernel. Both the Lawrence signal and the Evans designs each measure the maximum 670 points, while the new signals have the advantage of a far more even spacing and a lower maximum harmonic.

Comparing the two Evans designs, it is seen that harmonic vector (6-37) generates 36 points at which both second- and third-order kernel contributions fall. This problem is eliminated with the odd harmonic vector (6-39), at the cost of virtually doubling the maximum harmonic. If only the third-order kernel is to be measured the benefits of restricting the selection to odd harmonics could be questioned, since a sufficient number of distinct points can be measured using harmonic vector (6-37). However, using the odd harmonics of vector (6-39) will also ensure immunity from any fourth-order contributions.

The main restriction on the application of these signals is the increase in test time involved. For example, if harmonic vector (6-33) was required to cover the same bandwidth as a ten odd-harmonic multisine, its fundamental frequency would have to be

placed close to 13 times lower. For harmonic vector (6-39) the fundamental would have to be placed at around 125 times lower! This clearly implies that the signals are only suited to testing systems with fast dynamics.

TABLE 6-2

SUMMARY OF THIRD-ORDER SIGNAL DESIGNS

Eqn.	Signal	Harmonics			Kernel Measurement Points			
		Min	Max	$\sigma_{\Delta i}$	2nd-Order	Corrupt by 3rd	3rd-Order	Corrupt by 2nd
6-27	Lawrence 3rd	8	4099	684.8	100	7	670	7
6-29	Chua *et al.*	90	360	25.7	74	67	234	9
6-37	Evans 3rd	110	1187	8.0	100	36	670	36
6-39	Evans 3rd odd	241	2347	27.4	100	0	670	0

It is therefore useful to compare the signals in terms of test time per uncorrupted kernel measurement point, assuming the presence of third-order terms when measuring a second-order kernel and fourth-order terms when measuring the third. The values are shown in Table 6-3, normalised to the lowest value. It can be seen that similar values are obtained for each kernel with the Boyd and Chua signals and the new signal designs. This comparison does not take into account any settling time required and since the Boyd and Chua signal would have to be applied several times to measure the same number of points as the Evans signals, this gives the Evans signals an advantage.

TABLE 6-3

TIME PER KERNEL MEASUREMENT POINT

Second-Order		Third-Order	
Signal	Time / Point	Signal	Time / Point
Lawrence 2nd	1.4	Lawrence 3rd	5.4
Boyd *et al.*	1.1	Chua *et al.*	1.4
Evans 2nd	1.2	Evans 3rd	1.6
Evans 2nd odd	1	Evans 3rd odd	1.5

6.6 Crest Factor

Once the harmonic set has been chosen it is important to adjust the relative phases of these harmonics in order to minimise the signal CF. Reducing the CF allows the input harmonic amplitudes to be maximised for a given time-domain input range. Since the different output contributions fall at distinct frequencies and there is no interaction between them, their magnitude is directly proportional to the input amplitudes raised to the order of the nonlinearity. The phases of the input harmonics have no effect on the amplitude at each distinct point. For a given input amplitude constraint this means that amplitude measured at each kernel measurement frequency will be

$$P_{\text{meas}} \propto \frac{1}{\text{CF}^{p+2}} \tag{6-42}$$

As was previously discussed in Chapter 3, the CF minimisation formula proposed by Schroeder (1970) works well on signals with dense, evenly spaced spectra. However, it does not give good results on signals with sparse spectra, such as log-tones, as shown by Van der Ouderaa and Renneboog (1988). The $L\infty$ approach proposed by Guillaume *et al.* (1991) generally produces lower CFs than the Schroeder formula and other minimisation methods. A comparison of the CF obtained with zero, Schroeder and $L\infty$ phases for the signals described in Section 6.5 is presented in Table 6-4. It can be seen that the $L\infty$ algorithm produces a considerable improvement in the CF of each of the signals.

TABLE 6-4

SIGNAL CREST FACTORS

		Phases		
Eqn.	Signal	Zero	Schroeder	L∞
6-32	Evans 2nd	4.47	3.42	2.44
6-33	Evans 2nd odd	4.47	3.21	2.40
6-37	Evans 3rd	4.47	3.83	3.09
6-39	Evans 3rd odd	4.47	3.87	3.08

6.7 Application of Signals

The application of the signals designed in this chapter will now be illustrated using a simulation and two practical examples. The simulation will be used to illustrate the simultaneous measurement of the second- and third-order kernels of the nonlinear oscillation described by Duffing's equation. The first practical example is based on testing a "reference" nonlinear system composed of two linear systems, whose dynamics are known, and a multiplier. This serves to illustrate the application of the signals to practical systems, where problems such as dc offsets, nonlinear loading effects and measurement noise are encountered. In the second practical example, a signal is used to measure the second-order kernel of a dc servo-motor system.

6.7.1 Practical Considerations

A typical frequency-domain test and measurement set-up was shown in Figure 3-1, and there are a number of important considerations in its application to nonlinear systems.

Test Times. The increased test times associated with these signals have already been discussed, the second-order odd signal would involve a 13 fold increase in the signal period, when compared to a ten odd-harmonic signal. For the third-order odd signal the test time would be increased by 125 times. These increased test times are the price to be paid for gaining sufficient information about the Volterra kernels.

Number of Samples. The ratio of the highest harmonic frequency to the signal fundamental, along with the order of nonlinearity under test, determines the minimum number of samples per period. This can be expressed as

$$\text{Samples per period} \geq 2\,p\,\frac{f_{\max}}{f_0} \qquad (6-43)$$

which indicates that a large number of samples will have to be measured if the ratio of the maximum frequency to the fundamental is very high. This means a minimum of around 1,000 samples per period for the second-order odd signal of equation (6-33) and around 15,000 samples per period for the third-order odd signal of equation (6-39). This might have presented a problem in applying such signals in the past but the advent of fast VXI-based data logging systems, with large memories to store the signal samples, now makes their application a practical possibility.

Input Spectral Purity. The signal applied to the system must not have power at frequencies other than the input harmonics, since this will invalidate the basic assumptions of the testing method. To this end, a *reconstruction filter* should always be used to remove the higher harmonics introduced by the ZOH device. It is also important to eliminate any dc component at the system input, which may be introduced by a dc bias at the output of the ZOH or the reconstruction filter. Nonlinear loading effects may also generate power at unwanted input frequencies, and these must be eliminated using a recursive technique such as that proposed by Schoukens *et al.* (1993b), the application of which is described by Duym and Schoukens (1995).

Anti-Aliasing Filters. Care must be taken in the use of AA filters, since there is no longer a linear relationship between the components at a given frequency in the input and output. This means that it is no longer possible to eliminate the influence of the AA filters by means of a relative calibration of the measurement channels. From equation (6-24) and Figure 3-1, the estimated second-order kernel in the presence of AA filters can be expressed as

$$H_2(j\omega_1, j\omega_2) = \frac{Y(j\omega_1 + j\omega_2)\, G_o(j\omega_1 + j\omega_2)}{k\, U(j\omega_1)\, G_i(j\omega_1)\, U(j\omega_2)\, G_i(j\omega_2)} \qquad (6-44)$$

where $G_i(j\omega)$ and $G_o(j\omega)$ are the FRF's of the input and output filters. The following condition must hold in order that the filters do not introduce a bias into the estimate

$$\frac{G_o(j\omega_1 + j\omega_2)}{G_i(j\omega_1)\, G_i(j\omega_2)} = 1 \qquad (6-45)$$

which will only be met if the filters are ideal, with unity pass-band gain and linear phase. Since such filters cannot be realised in practice, there are three options:

(1) Carry out an absolute calibration of each measurement channel and compensate the measured data;

(2) Set the bandwidth of the AA filters carefully to minimise their influence on the measured input and output, accepting that small errors will be introduced;

(3) Dispense with AA filters altogether, while relying on the reconstruction filter to remove the higher harmonics generated by the ZOH and sampling faster to counter the increased influence of noise.

The third option was employed in the practical examples presented in this chapter.

6.7.2 Simulation of Duffing's Equation

Duffing's equation is used to describe a wide class of nonlinear oscillations in engineering applications. The example used is taken from Zhang and Billings (1993), where it was used to illustrate the derivation of second- and third-order FRFs from a NARMAX model. The equation of motion is given by

$$\frac{d^2y(t)}{dt^2} + 2\zeta\omega_n \frac{dy(t)}{dt} + \omega_n^2 y(t) + a_2\omega_n^2 y(t)^2 + a_3\omega_n^2 y(t)^3 = u(t) \qquad (6-46)$$

where ζ is the damping factor and ω_n is the undamped natural frequency (rad/s) and a_2 and a_3 are the coefficients of the second- and third-order nonlinear terms. The coefficient values used by Zhang and Billings were $\zeta = 0.1$, $\omega_n = 100$ rad/s, $a_2 = 10^3$ and $a_3 = 5 \times 10^5$. The Evans third-order odd signal was used, with the harmonic vector of equation (6-38), a fundamental of 0.01 Hz and a reconstruction frequency of 500 Hz.

The measured second-order kernel is shown in Figure 6-12, where its resonant nature is clearly visible. The missing values along the $f_1 = -f_2$ diagonal are the result of the unmeasured Type I contributions. A more complete view of the kernel can be obtained by viewing the surface from above and using colour as an amplitude, which is illustrated in Figure 6-13. The resonant ridge in the kernel is clearly visible in this plot. The measured third-order kernel is shown in Figure 6-14 plotted as a cube with colour as an amplitude and one corner removed. Once again, the missing values are those corresponding with Type I contributions and the resonant nature of the system is clearly visible.

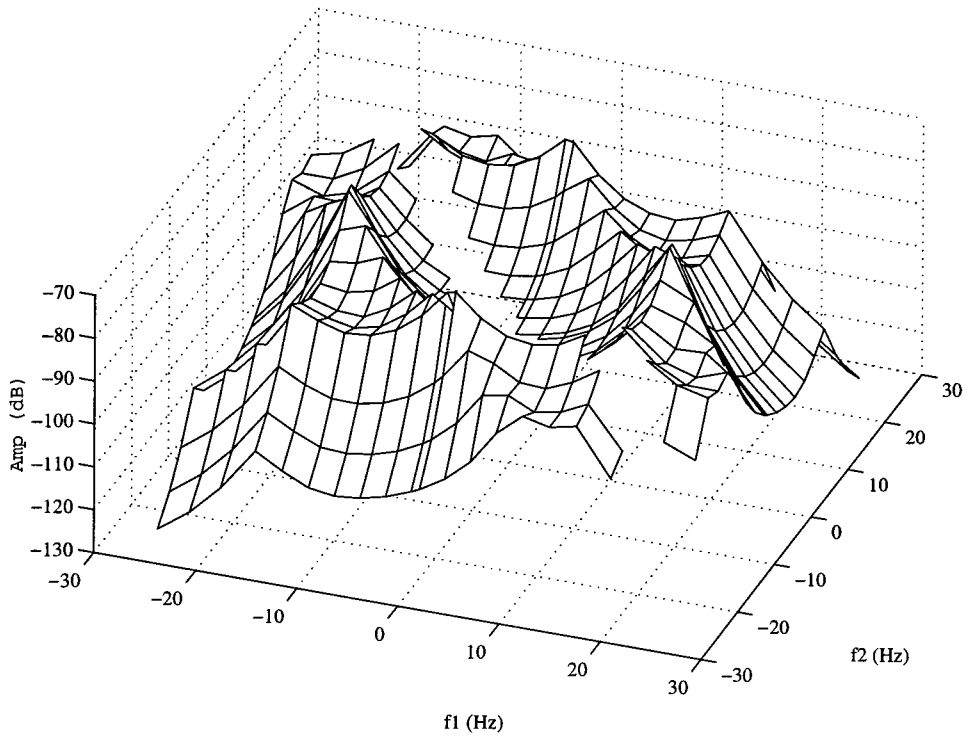

Figure 6-12. Amplitude of the second-order kernel of Duffing's equation.

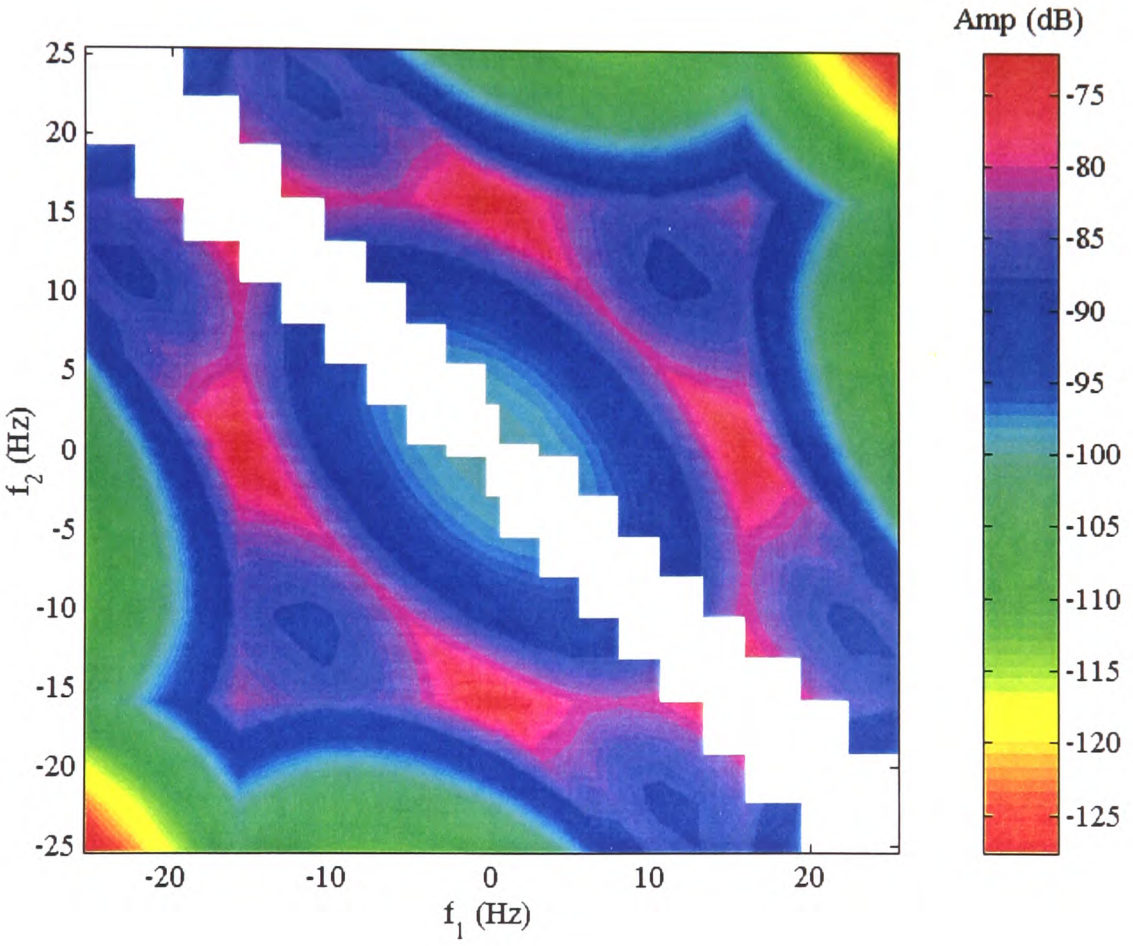

Figure 6-13. Pseudo-colour plot of the second-order kernel amplitude.

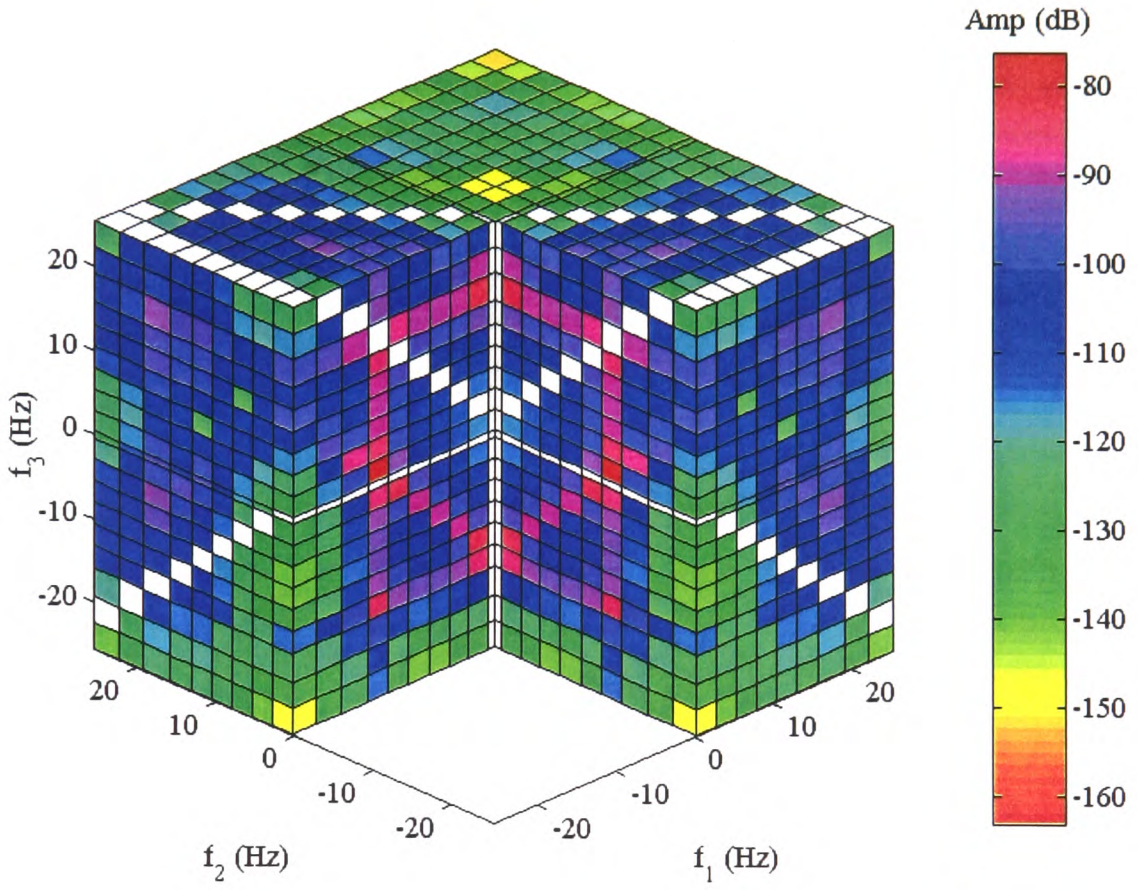

Figure 6-14. Amplitude of the third-order kernel of Duffing's equation.

6.7.3 Reference Nonlinear System

In order to assess the practical performance of the signals a reference nonlinear system was constructed, of the form shown in Figure 6-1. The linear elements consisted of two electrical circuits, these being a first-order low-pass system and a second-order low-pass system with resonance. The first step was to identify s-domain models of each of the linear subsystems, using the frequency-domain estimator presented in Chapter 2. Volterra algebra was then used to calculate the symmetrised second-order kernel of the overall system, using equation (6-6). The calculated kernel was then used as a reference for the practically measured kernel values, which allowed the influence of various nonidealities encountered during practical testing to be assessed.

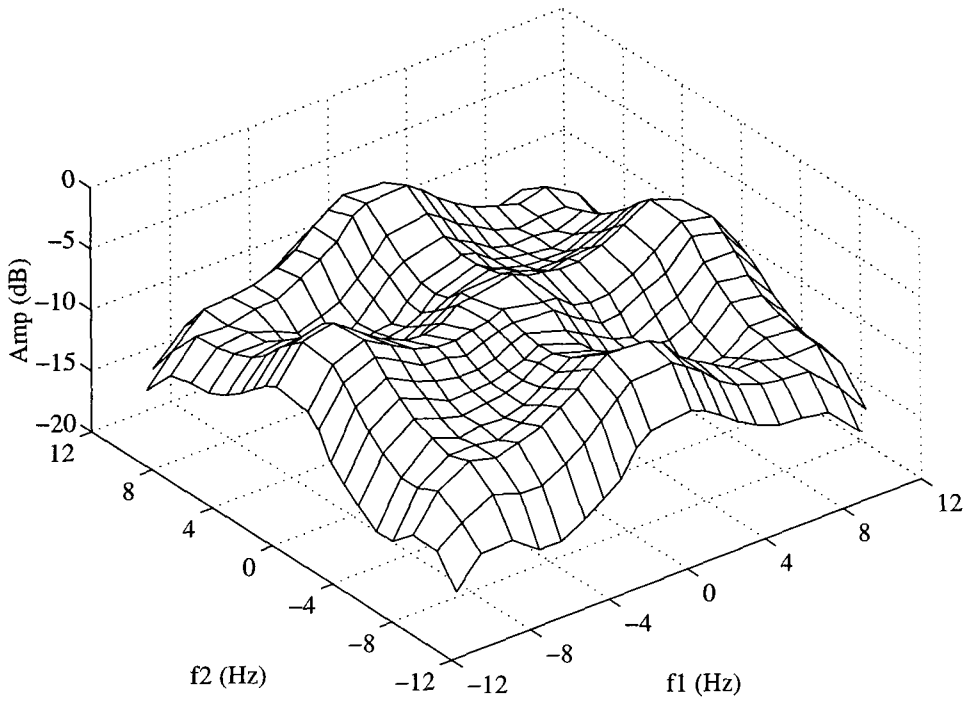

Figure 6-15. Measured amplitude of second-order kernel.

Measurements of the overall system were then carried out using the Evans second-order signal, with the harmonics given in equation (6-32). The test conditions were as follows:

- $f_0 = 0.085$ Hz, $f_{max} = 11.65$ Hz, $f_s = 200$ Hz, Reconstruction filter at 30 Hz, AA filters omitted. Periods measured (M) = 6.

The result of a single test is shown in Figure 6-15, where the amplitude of the measured second-order kernel is plotted. In this case a crude "nearest neighbour" interpolation technique was used to fill in the values along the $f_1 = -f_2$ diagonal, in order to illustrate the problem with the missing measurements. The absolute relative error between the measured and reference kernels is shown as a percentage in Figure 6-16. The error is seen to increase at higher frequencies and also to be large along the interpolated diagonal. This error could be reduced by using the signal of paired harmonics defined in equation (6-35) or by using an improved interpolation technique.

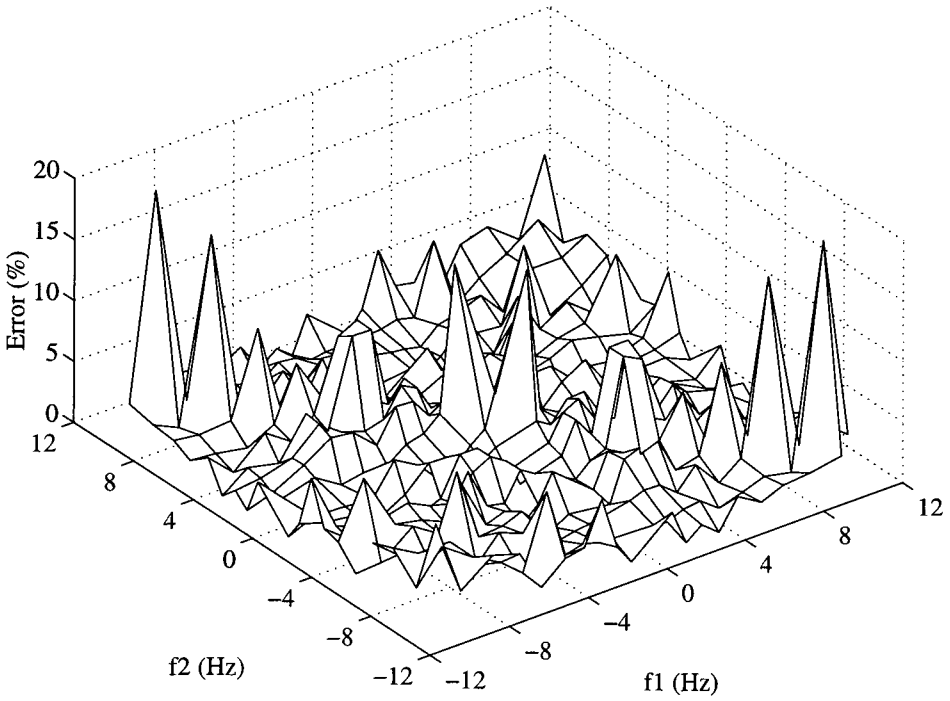

Figure 6-16. Relative error in kernel measurement.

6.7.4 Servo Motor System

The second practical example involves the testing of a dc servo-motor system, the schematic diagram for which is shown in Figure 6-17. These are standard components of the MS150 modular servo system manufactured by Feedback Instruments, which is extensively used for teaching purposes. The motor was tested in its armature-controlled configuration, by varying the forward speed only, in order to avoid the dead-zone nonlinearity known to be present around the zero crossing.

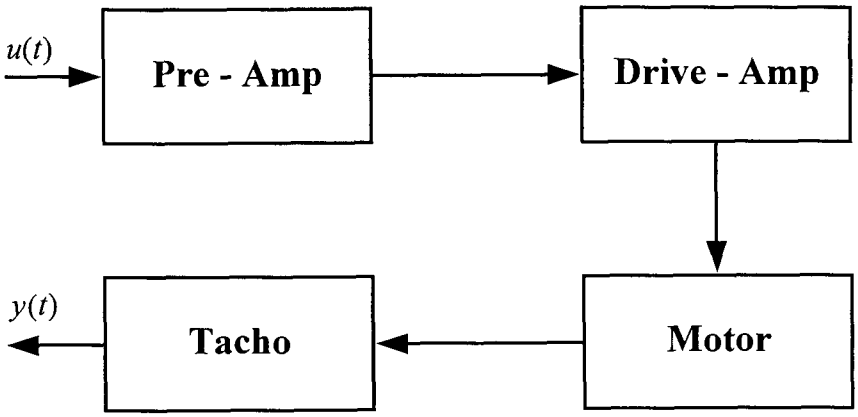

Figure 6-17. Schematic diagram of servo motor system.

A pilot test using a odd-odd multisine provided initial information about the bandwidth of the system and revealed the presence of a significant even-order nonlinearity, along with some lower level odd-order effects, as shown in Figure 6-18. Ten periods of the signal were measured and the periodicity of the nonlinear contributions was assessed using the nonlinear coherence function, which is plotted in Figure 6-19. It is clear from this plot that a significant even-order nonlinearity is present in the system.

The system was then tested using a signal with the harmonic vector of equation (6-33), in order to ensure that the second-order kernel measurements would be unaffected by the output of any odd-order terms. The test conditions were:

- $f_0 = 0.03$ Hz, $f_{max} = 6.81$ Hz, $f_s = 150$ Hz, Reconstruction filter at 20 Hz, AA filters omitted. Periods measured (M) = 6.

The input and output spectra of the servo motor are shown in Figure 6-20, where the additional second-order contributions are clearly visible at the output.

The nonlinear coherence is plotted in Figure 6-21, with the coherence of the *kernel measurement points* shown green, the input frequencies shown red and all other frequencies shown blue. This convention is deliberately different from that used in previous plots, in order to focus attention on the kernel measurement frequencies themselves. The plots show that they have a low coherence at the input and a high coherence at the output. The output coherence of every kernel measurement frequency is greater than 0.78, indicating the presence of periodic second-order nonlinear contributions.

A plot of the measured kernel amplitude is shown in Figure 6-22, with the missing $f_1 = -f_2$ values calculated using cubic spline interpolation. The plot shows that the kernel exhibits a low-pass nature and that it is chiefly characterised by values along the $f_1 = f_2$ diagonal. This can be seen more clearly in the pseudo-colour plot of the kernel shown in Figure 6-23. This particular form of Volterra kernel is associated with the Hammerstein nonlinear model shown in Figure 4-1 (b), where a static nonlinearity precedes a dynamic linear element (Weiss *et al.*, 1995). The dominant linear dynamics in this frequency range are clearly those of the motor itself, which can be modelled by a mechanical time constant of around 1.4 s. This suggests that the nonlinearity precedes the linear dynamics of the motor and is to be found in one of the drive amplifiers.

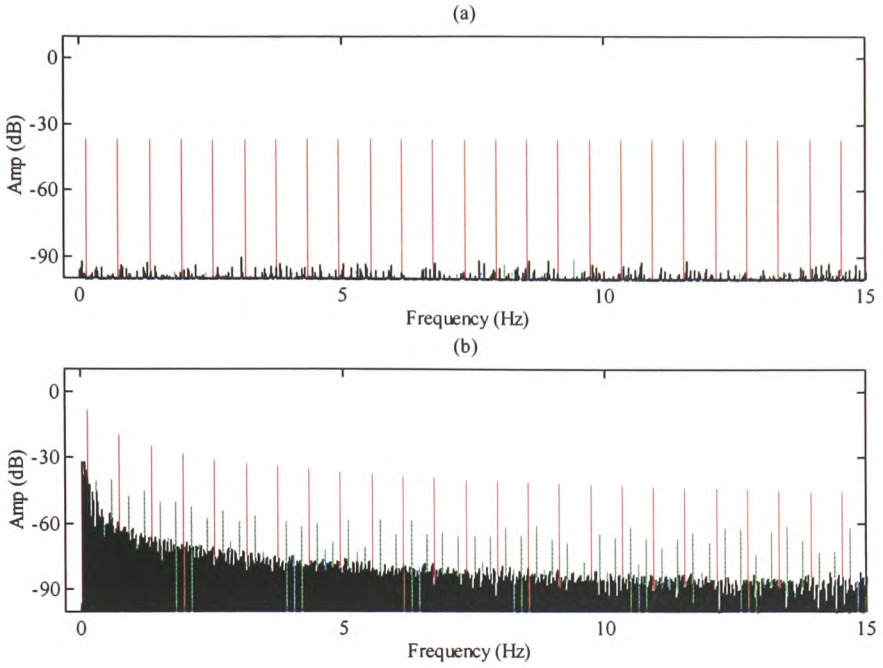

Figure 6-18. Spectrum of the pilot test signal at the (a) input and (b) output. Input harmonics (red), excluded even harmonics (green), excluded odd (blue). All other frequencies shown black.

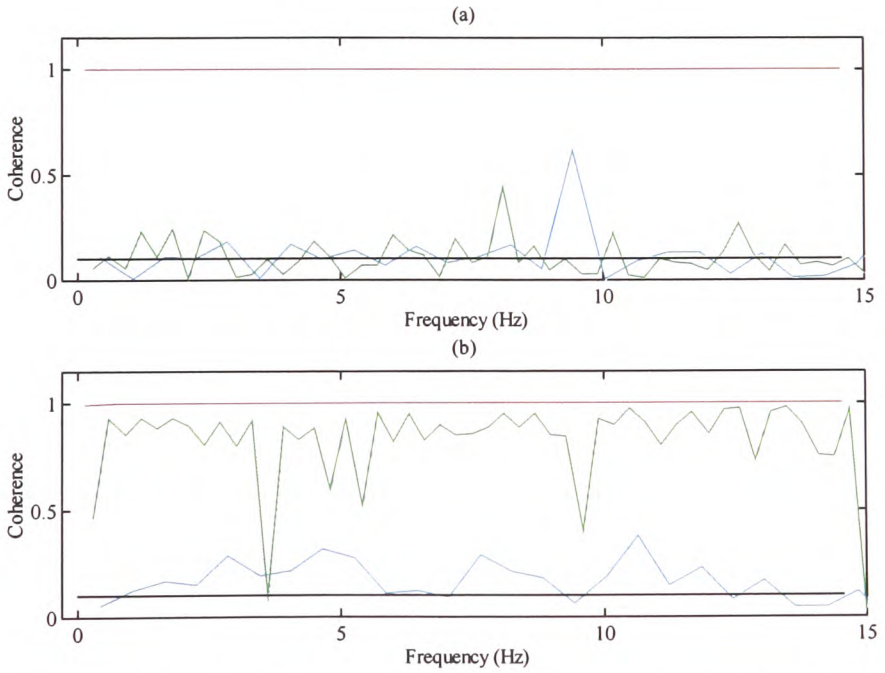

Figure 6-19. Nonlinear coherence of the pilot test signal at the (a) input and (b) output. Input harmonics (red), excluded even harmonics (green), excluded odd (blue). With $1/M$ bound shown black.

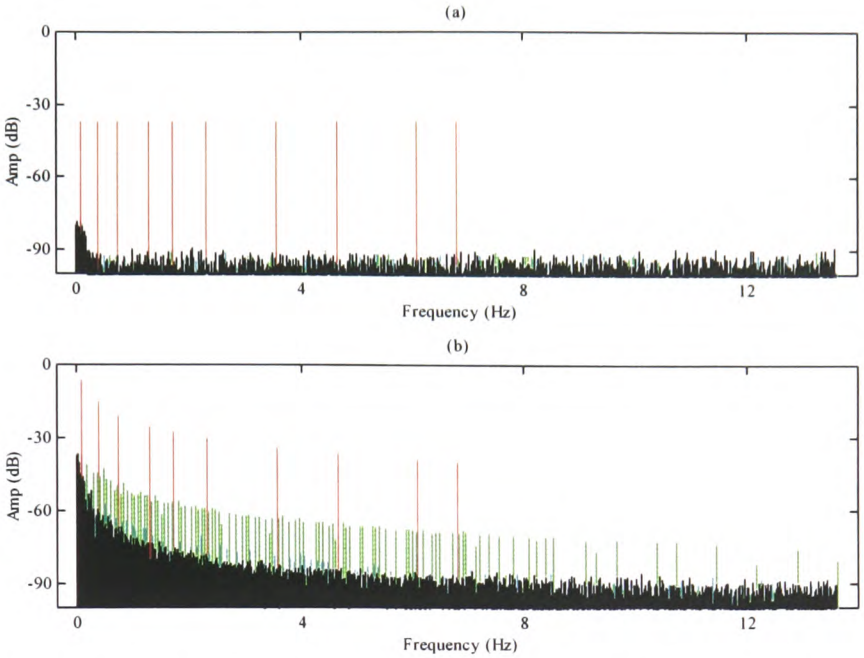

Figure 6-20. Spectrum of the Evans odd signal at the (a) input and (b) output. Input harmonics (red), excluded even harmonics (green), excluded odd (blue). All other frequencies shown black.

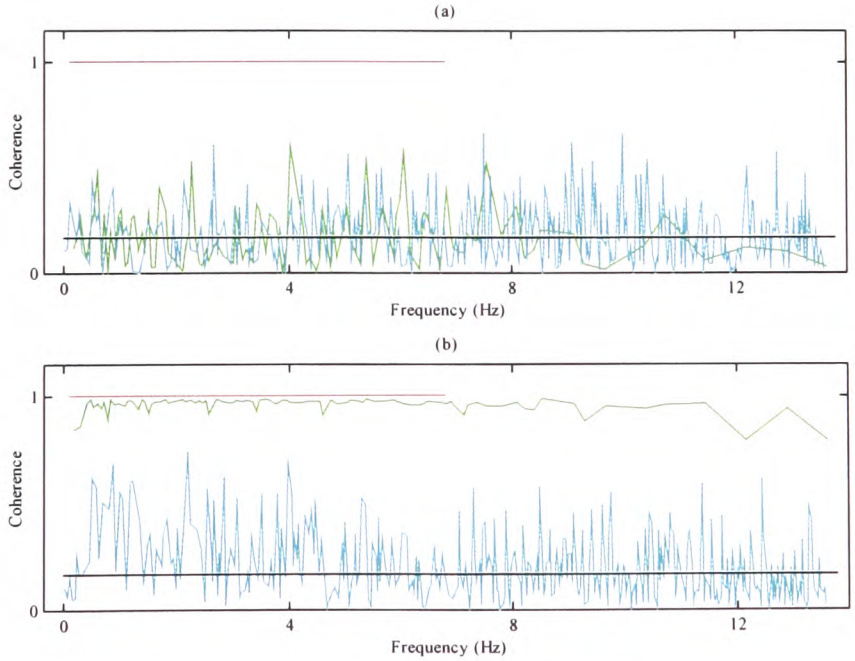

Figure 6-21. Nonlinear coherence of the Evans odd signal at the (a) input and (b) output. Input harmonics (red), kernel measurement points (green), all other harmonics (blue). With $1/M$ bound shown black.

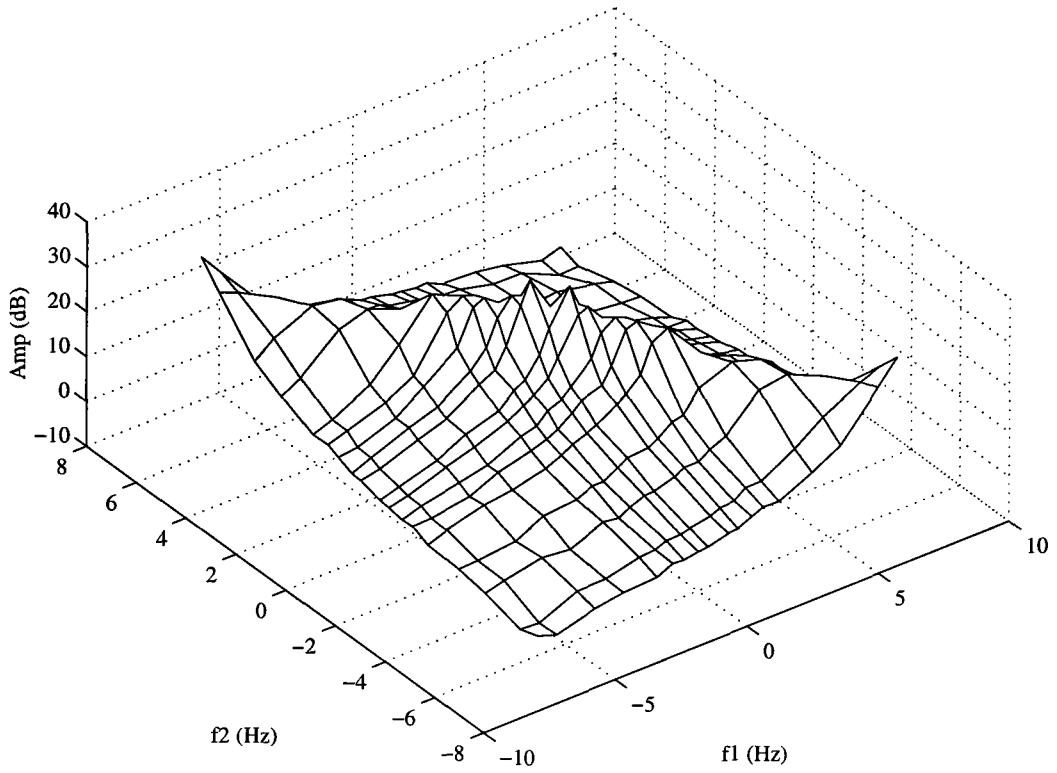

Figure 6-22. Amplitude of servo motor second-order kernel.

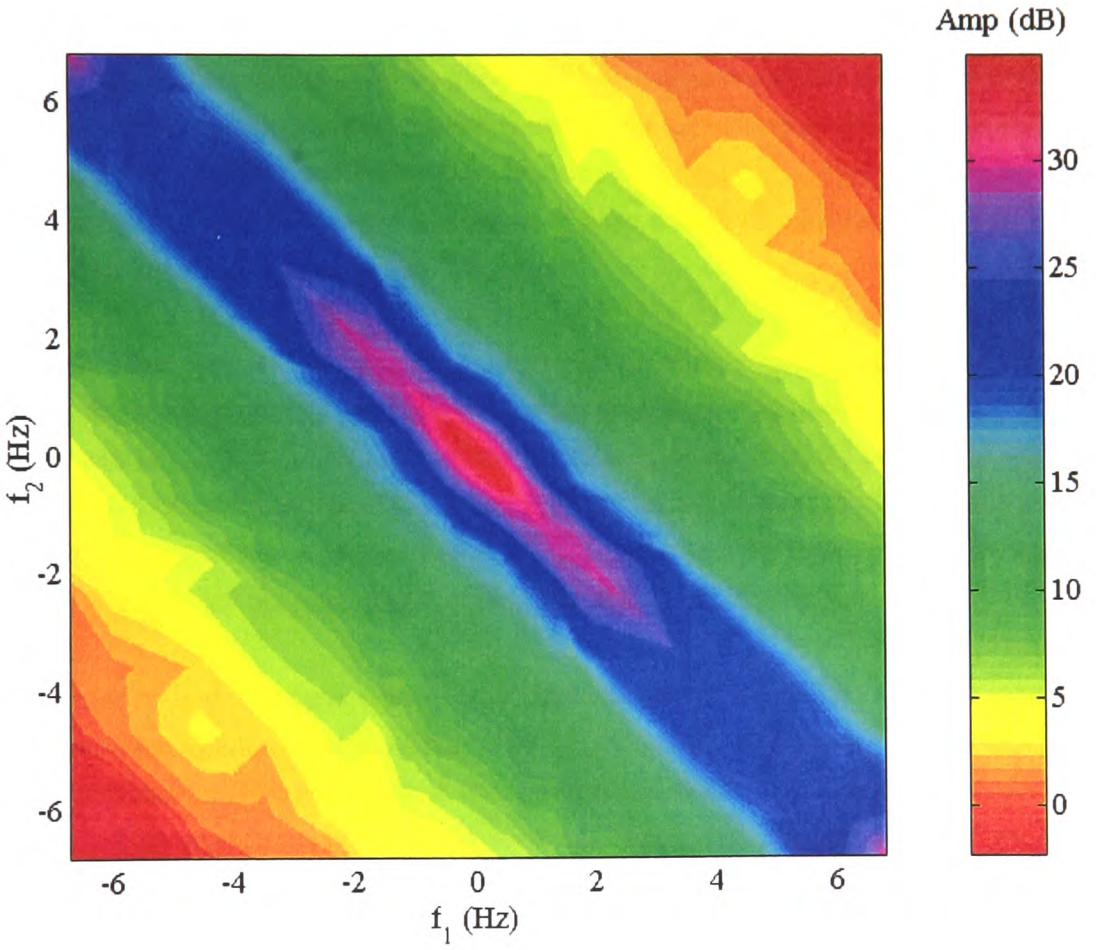

Figure 6-23. Pseudo-colour plot of the servo motor second-order kernel amplitude.

6.8 Conclusions

A review has been presented of the Volterra representation of nonlinear systems, with particular emphasis on the harmonic output of such systems when driven by periodic multifrequency inputs. This has enabled a range of periodic signals to be defined, for directly measuring points on the HOFRF of Volterra kernels. A study has been made of previously designed signals of this type and their drawbacks discussed. The new designs overcome some of these problems, by allowing the maximum possible kernel points to be measured and maintaining a near-even harmonic spacing. A straightforward procedure for constructing suitable harmonic vectors has also been described. Harmonic vectors for measuring second-and third-order kernels have been presented and the measurement of higher-order kernels has been briefly discussed.

A problem associated with all such signals has also been highlighted, in that points on the kernel which correspond to Type I output contributions cannot be directly measured. This problem can be partially overcome by designing signals made up of paired harmonics, which allow measurements close to the unmeasured values. The other drawback of these signals is the significant increase in test times associated with their application, which suggests that they are only suitable for testing systems with fast dynamics.

The minimisation of the signal crest factors has been examined, and the application of the signals has been illustrated in simulation and by two practical examples. The simulation study examined the measurement of the second and third-order kernels of Duffing's equation. This replicated an example used by Zhang and Billings to illustrate how HOFRF's can be obtained using the NARMAX approach. The simulation shows how the signals designed in this chapter can be used to directly measure point on the HOFRF's and hence validate the HOFRF's obtained using other methods.

The practical problems associated with the testing of nonlinear systems have been highlighted. Practical results were presented for a reference nonlinearity and a servo motor system. The reference nonlinearity consisted of two known linear systems and a multiplier. This allowed the measured kernel values estimated using the signals to be compared with the true kernel values, calculated using Volterra algebra. The second example dealt with the measurement of the second-order kernel of a servo-motor system. The measured kernel was found to have characteristics associated with a Hammerstein model, which suggested that the second-order nonlinearity was present in the amplifier sections of the system.

The periodic signals designed in this chapter offer an attractive alternative to the use of random noise signals for the measurement of Volterra kernels and are a significant improvement on previously designed signals of this type.

Chapter VII

Gas Turbines

Abstract — *This chapter serves to introduce the dynamic modelling of gas turbines, beginning with an outline of their construction and operation. This is followed by a discussion of the basic requirements of a good engine model. A review of the work previously conducted in this field is then presented, which examines both the models derived and the engine testing methods employed. Attention is paid to modelling the basic dynamic response of the engine and to modelling thermal effects during large transient manoeuvres. The limitations of previous work are highlighted and the overall aims of the present work are defined.*

7.1 Introduction

The modelling of gas turbines has been the subject of considerable study since the early days of jet propulsion. The development of aircraft gas turbines can be traced back to the 1930's, with the simultaneous development of jet engine technology in Britain and Germany. The first operational jet engine was built by a British team headed by Frank Whittle and tested in April 1937. This led to the development of an experimental aircraft, the Gloster E28/29, which first flew in May 1941. Meanwhile, Hans von Ohain and Max Hahn were developing a similar engine for the Heinkel aircraft company and the first flight of the He-178 took place in August 1939. A comprehensive overview of gas turbine development has been presented by Ohain (1996).

Gas turbine engines are now extensively used in aero, marine and industrial applications. There is also considerable cross-fertilisation between steam and gas turbine technologies, such as the collaboration between Westinghouse and Rolls Royce, resulting in the development of small bespoke power stations based on aero-engines (IEE Review, 1994). Gas turbines are even finding application in a new generation of low-emission electric cars, currently being tested by Volvo (Vernet, 1994).

This chapter serves to introduce the concepts of gas turbine modelling and review the work conducted in this area to date. Chapters 8 and 9 will present the results of testing and modelling a Rolls Royce Spey engine using multisine excitations and frequency-domain estimation techniques.

7.2 The Gas Turbine

A gas turbine is made up of three basic components: a compressor, a combustion chamber and a turbine. Air is drawn into the engine inlet by the compressor, which compresses it and then delivers it to the combustion chamber. Within the combustion chamber the air is mixed with fuel and the mixture ignited, producing a rise in temperature and hence an expansion of the gases. These gases are exhausted through the engine nozzle but first pass through the turbine, which is designed to extract sufficient energy from them to keep the compressor rotating, so that the engine is self sustaining.

The simplest form of gas turbine is known as a *turbojet*, where all the intake air passes through the engine core. Such engines have a small frontal area and high jet velocity, making them suitable for use in high speed aircraft. The Rolls Royce Spey Mk 202 tested

in this work is an example of a *turbofan* engine, which is currently the most common type of gas turbine used for aircraft propulsion. The Spey is a comparatively small engine by modern standards and can produce about 55 kN of thrust without reheat. Although it is no longer in service it has the same basic architecture, for purposes of control, as that of more modern engines such as the EJ200, which is used to power the Eurofighter (Dadd *et al.*, 1996).

A simplified diagram of a Spey engine is shown in Figure 7-1, where it can be seen that both the compressor and the turbine are split into *low pressure* (LP) and *high pressure* (HP) stages. The HP turbine drives the HP compressor and the LP turbine the LP compressor. They are connected by concentric shafts which rotate at different speeds, which will be denoted N_H and N_L. Each combination of compressor, shaft and turbine is called a *spool.* The twin-spool design was developed in order to improve engine efficiency by increasing the overall compressor pressure ratio. This arrangement reduces the pressure ratio required in each compressor stage, thus simplifying the aerodynamic design.

It can also be seen that only part of the air entering the engine passes through the HP compressor and the combustion chamber, while the remainder by-passes the engine core. The ratio between the mass flow of air in the by-pass duct and in the core is termed the *by-pass ratio.* The Spey has a by-pass ratio of only 0.6, while many modern commercial engines, such as the Rolls Royce Trent, have by-pass ratios as high as 4.8. The by-pass air reduces the overall jet velocity, which leads to lower noise levels, along with better efficiency and reduced fuel consumption.

The Spey engine also incorporates a number of additional features associated with modern military gas turbines. Variable *inlet guide vanes* (VIGV's) control the angle at which the air flows into the HP compressor and these are adjusted in relation to the engine speed, in order to improve engine performance. A variable area *reheat nozzle* at the engine output allows additional fuel to be injected into the hot exhaust gases, to produce additional thrust when required. The nozzle area is variable in order to allow rapid acceleration and deceleration of the engine, without causing it to surge. Finally, an adjustable HP *compressor bleed valve* allows the mass flow through the later compressor stages to be reduced, in order to lower the pressure ratio across the engine and prevent it from going into surge at low speeds.

A *surge* condition is reached when the pressure difference across a compressor is too great, for a given engine speed and mass flow, causing a violent breakdown of air-flow within that compressor. The hot high-pressure gases in the engine core can then reverse their direction of motion and surge backwards through the compressor. A related malfunction which must also be avoided is known as *deep stall*, which is itself a severe case of *rotating stall*. This involves a stable and sustainable reduction of the mass flow and the pressure ratio across the compressor, which can result in the turbine overheating.

The shaft speeds are the primary outputs of a gas turbine, from which the internal engine pressures and thrust can be calculated. The dynamic relationship between these shaft speeds and the engine fuel feed will be studied in the following chapters.

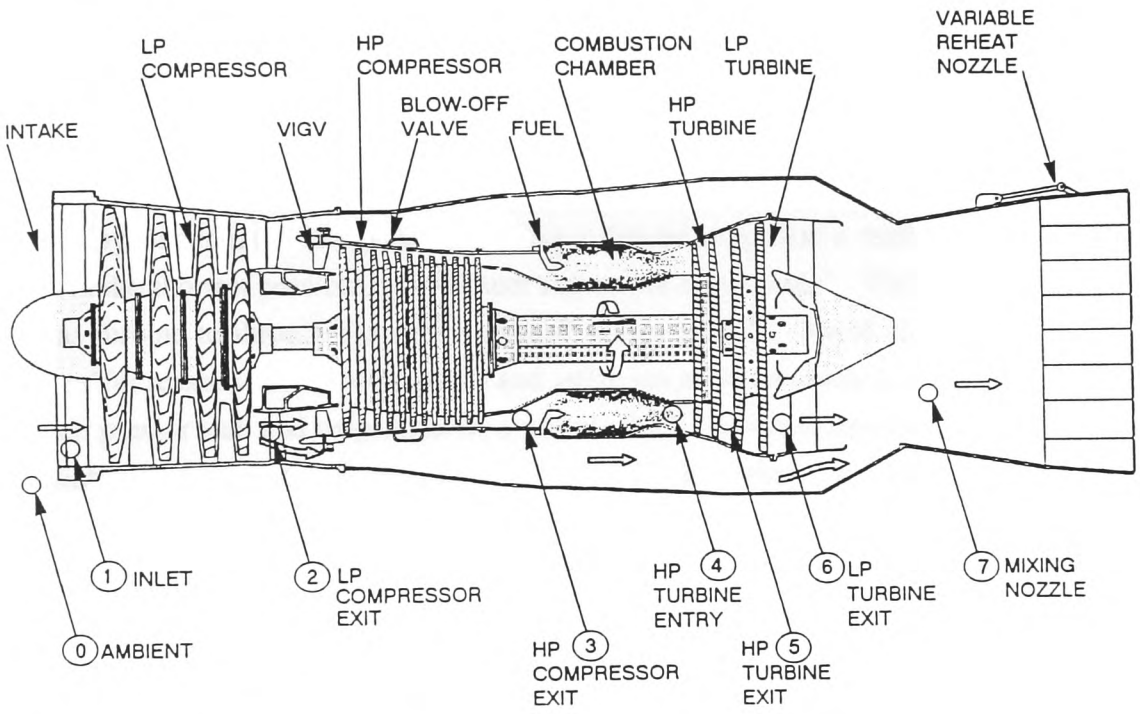

Figure 7-1. Simplified schematic of a Rolls Royce Spey engine, showing main engine components and station numbers.

7.3 Modelling Gas Turbines

Engine models are required both in the development and operational stages of the life of a gas turbine. In the development stage the models can provide insights into the engine behaviour and physics and allow the overall control system specifications to be defined. This requires that comprehensive models are developed, which incorporate all the secondary effects encountered in the engine. Such detailed knowledge is of increasing importance, since it is now common practice for companies purchasing gas turbines to insist on guaranteed engine response characteristics within their contracts.

While the modelling effort may well be expensive, it should yield significant savings in terms of the overall engine development. The performance of any engine model must then be verified using real data gathered from the gas turbine, after which the model parameters may need to be updated. A physical interpretation can then be made of these model parameters, in order to verify assumptions about the engine characteristics.

An important area for the application of such models is that of *engine handling*, which is defined by Bauerfeind (1982) as: "achieving a desired state with a minimum of manual effort in the shortest possible time without any undue safety risks". The response rate of the gas turbine to throttle changes is critical in this respect. Rapid response rates are required for reasons of manoeuvrability and safety but problems such as shaft over-speeds and compressor stall and surge must be avoided. Such considerations are obviously vital for military aircraft but are also important in the civilian sphere, to cope with situations such as a baulked landing, where a landing is aborted at the last minute and a rapid engine response is required.

The range of different throttle changes which must be dealt with by a given engine and its control system are illustrated in Figure 7-2, which shows the throttle movement patterns for the leader and wingman of the Red Arrows aerobatic team. It can be seen that the leader uses very little of the throttle during the main part of the flight, while the wingman uses nearly all of the throttle range. The design of an engine and control system which can cope with these differing requirements is greatly facilitated by the use of model based simulations at the development stage.

The use of models in the operational stage of an engine's life is usually restricted to functions such as controller tuning, the training and familiarisation of pilots, condition

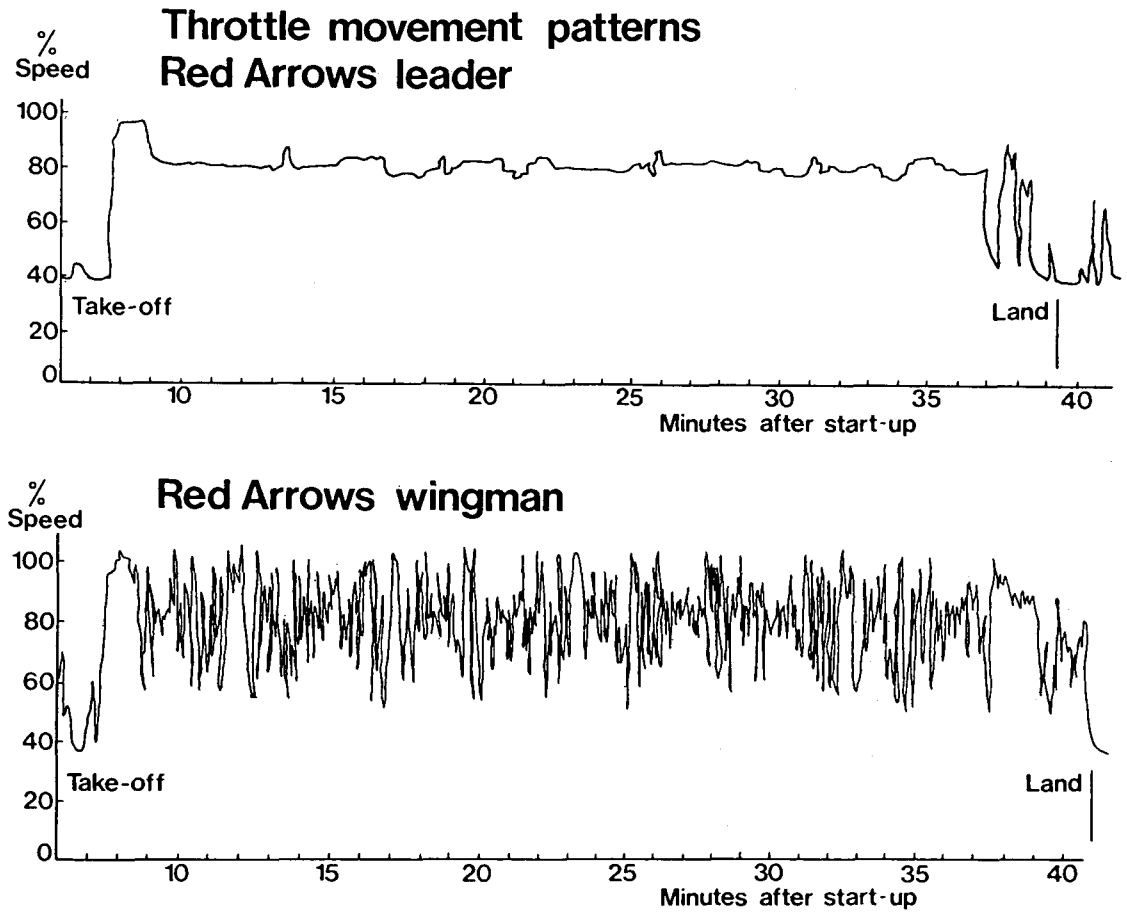

Figure 7-2. Throttle movement patterns for Red Arrows leader and wingman. (Reproduced from Pollitt (1982), with permission.)

monitoring and fault diagnosis (Zhang *et al.* (1994) and Milne and Travé-Massuyès (1994)). The models used for these applications are generally less sophisticated than those used in the development stage.

The requirements for a good engine model were discussed by Mats and Tunakov (1982) and have been summarised by Saravanamuttoo and MacIsaac (1982), the three most important being:

(1) Flexibility: the model must be capable of handling all the obvious flight scenarios and be flexible enough to cope with situations which were not initially anticipated. It must also be possible to update the model in a straightforward manner to reflect any changes in engine and control system design.

(2) Credibility: the model output should be in a form which is readily understandable to performance and development engineers and should generate data which correspond to the outputs obtained from real engine tests.

(3) Reliability: the model must be capable of being easily checked for correct functioning; this is especially true for complex engine simulations.

The need for flexibility will tend to push towards more complex models, a development which Bauerfeind (1982) has cautioned against, arguing that: "Although the most important physical effects governing engine transients are well established and very comprehensive computer models including these effects are available in industry there appears to be a need for simplified models for easy handling with a more transparent approximation of these physical effects."

The work presented in this thesis is particularly concerned with establishing credible and reliable models, which are at the same time transparent in terms of the physical effects which they are modelling.

7.4 Previous Work

A review of the history of engine modelling will now be presented, in order to highlight the problems which have yet to be overcome. The earliest engine models were derived by Gold and Rosenzweig (1952), who deduced that the fuel feed to shaft speed dynamics of a single shaft engine could be modelled as a first-order system. The time constant was expressed in terms of partial derivatives, which made it difficult to interpret.

In work conducted for the Lucas company, Lawrence and Powell (1957) developed an analysis based on the assumption that a sudden increase in fuel flow produced an instantaneous increase in turbine torque but zero increase in compressor torque. This greatly simplified the analysis, allowing the rotor time constant of a single shaft engine to be expressed in terms of thermodynamic and mechanical parameters

$$\tau = k\, I\, N \left(\frac{dW_f}{dN} \frac{\Delta T_{34}}{T_3} \right)^{-1} \qquad (7-1)$$

where k is a dimensionless constant, I the polar moment of inertia of the rotor, N the rotational speed and W_f the fuel flow. T_3 is the turbine inlet temperature and ΔT_{34} the

temperature drop across the turbine. Figure 7-3 shows the typical variation of $\Delta T_{34} / T_3$ and dW_f / dN for a simple turbojet. Both terms increase with increased speed, causing the shaft time constant to become smaller. Figure 7-4 shows the theoretically predicted variation, along with some practical results obtained by Lawrence and Powell using step response and single sine tests.

Lawrence and Powell also studied twin-shaft turbojets and concluded that each shaft could be modelled by a first-order transfer function. This was theoretically justified by the assumption of tight *aerodynamic coupling* between the shafts. This assumes that the HP shaft can be considered as equivalent to a single spool engine with a fixed nozzle area, the value of which is determined by the throat area of the LP turbine stator. If the LP turbine is operating at the maximum possible mass flow for a given LP turbine inlet pressure and temperature, a condition known as *choked flow*, then the HP turbine operating line will be uniquely defined. It is known that the LP turbine will be choked for most of the useful operating range. This means that once the LP shaft reaches a given operating point the HP shaft operating point will also be fixed (Cohen *et al.*, 1987, pp. 322-323).

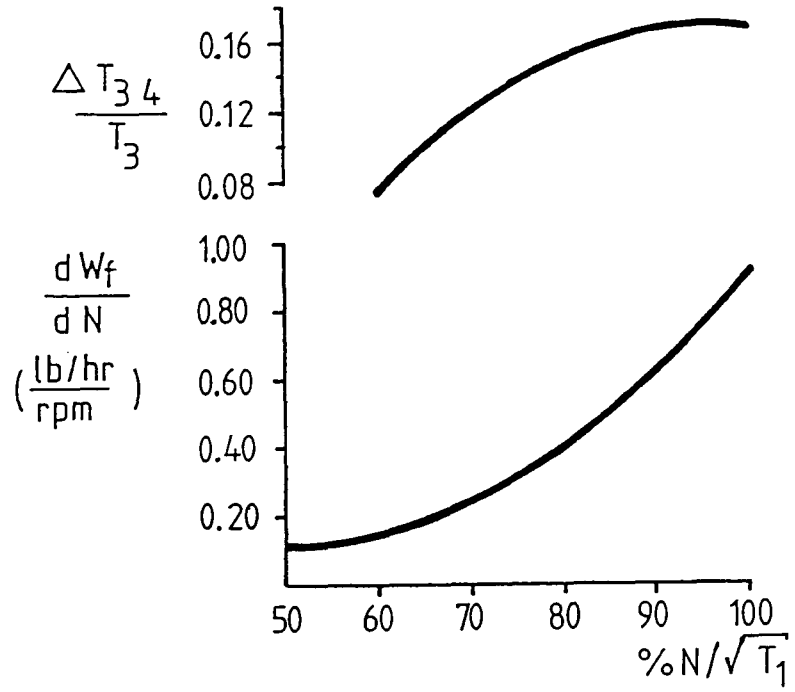

Figure 7-3. Parameters determining the shaft time constant.
(Reproduced from Saravanamuttoo (1992), with permission.)

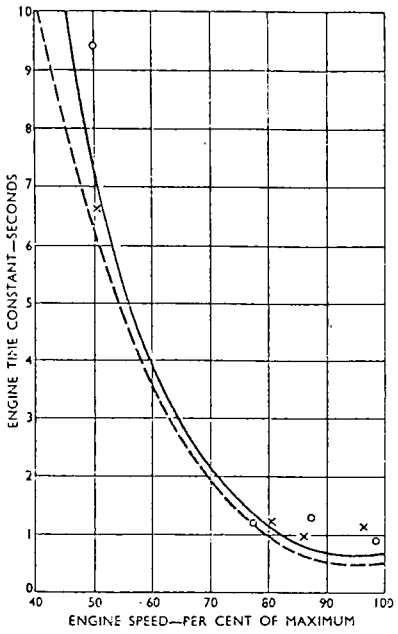

Figure 7-4. Comparison of theoretical and practical results on a Sapphire engine. Theoretical (solid), approximate theoretical (dashed), step tests (crosses), single sine tests (circles). (Reproduced from Lawrence and Powell (1957), with permission.)

Work of a similar nature was conducted at the National Gas Turbine Establishment (NGTE) by Fitchie *et al.* (1959), who derived one zero, two pole models for the fuel feed to shaft speed dynamics of the Rolls Royce Olympus 101, a twin-shaft turbojet without by-pass air. The model time constants were expressed in terms of partial derivatives of the HP and LP shaft torques with respect to the fuel feed and the shaft speeds. The model structures derived in this way imposed common poles on the HP and LP shafts, while allowing the values of the HP and LP zeros to differ.

The partial derivatives were evaluated experimentally and it was found that the partial derivative of the HP shaft torque with respect to the LP shaft speed was very small. This resulted in HP shaft models with cancelling pole-zero pairs, which could be reduced to first-order. The equivalent term in the LP shaft model was more significant, resulting in second-order models with a dominant time constant equal to that of the HP shaft and a weak second-order effect. This is illustrated in Figure 7-5, which shows the pole and zero

values of the shaft models plotted for a range of operating points. There is a clear pole-zero cancellation in the HP shaft model at all operating points, while the poles and zeros are more distinct in the LP shaft model, particularly at higher operating points.

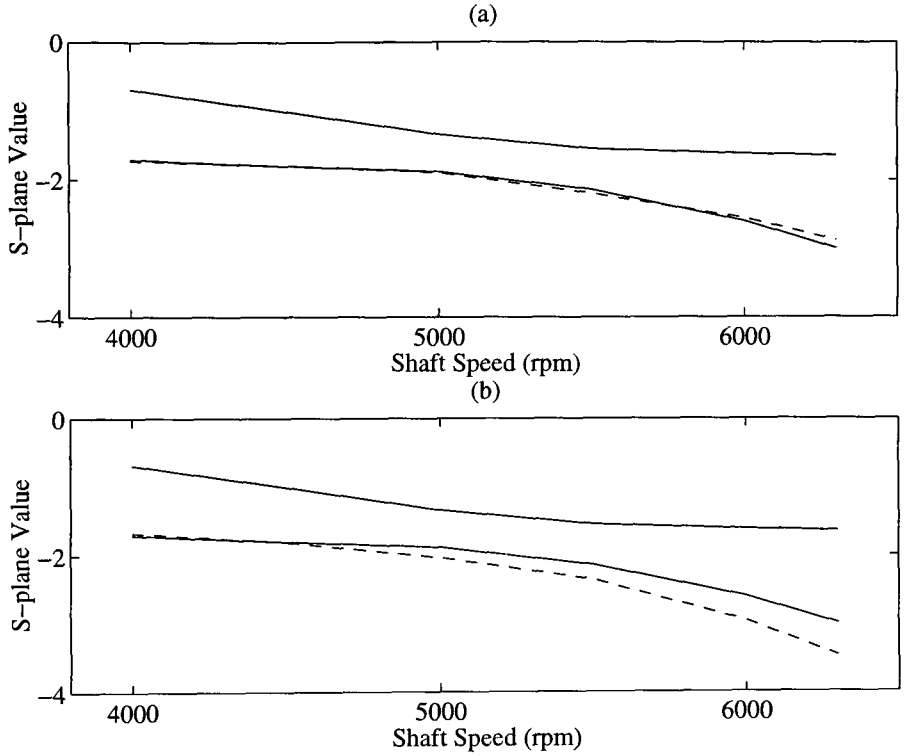

Figure 7-5. Poles and zeros of Fitchie *et al.* models for Olympus engine (a) HP shaft and (b) LP shaft. Showing poles (solid) and zeros (dashed).

The derived model structures, along with the numerical values obtained at an LP shaft speed of 6000 rpm, are shown in equations (7-2) and (7-3). Expressing the transfer functions in partial fraction form makes clear the relative contribution of each pole to the shaft response.

$$H_{HP}(s) = \frac{(\tau_{HP}s + 1)}{(\tau_{N_H}s + 1)(\tau_{N_L}s + 1)} = \frac{(0.391s + 1)}{(0.619s + 1)(0.384s + 1)} = \frac{1.567}{(s + 1.615)} + \frac{0.0776}{(s + 2.604)} \tag{7-2}$$

$$H_{LP}(s) = \frac{(\tau_{LP}s + 1)}{(\tau_{N_H}s + 1)(\tau_{N_L}s + 1)} = \frac{(0.338s + 1)}{(0.619s + 1)(0.384s + 1)} = \frac{1.932}{(s + 1.615)} - \frac{0.510}{(s + 2.604)} \tag{7-3}$$

These models were verified by extensive engine testing, using step and single sine inputs at a range of engine operating points, whose amplitudes were restricted to ± 7% of the steady-state W_f (henceforth ± 7% W_f). The results showed that each shaft could be modelled as first-order, with almost identical time constants. It was not possible to identify the second-order dynamics of the LP shaft, within the limits of the available measurement and modelling techniques. The work also confirmed the variation of the time constants with engine speed. No effort was made to assess the degree of uncertainty associated with the frequency response measurements or the models fitted to them.

Both Lawrence and Powell (1957) and Fitchie *et al.* (1959) focussed on turbojet engines, in which all of the intake air passes through the engine core. This means that the mass flow through the engine is fixed by one choked point, the LP turbine inlet, and a tight aerodynamic coupling exists between the shafts. This is not the case with turbofan engines, where a certain portion of the intake air by-passes the core. The total mass flow is no longer fixed by one choked point and this introduces an additional degree of freedom into the engine response. In discussing the response of a twin-shaft turbofan to a change in fuel feed, Saravanamuttoo and MacIsaac (1982) argued that while the behaviour of the HP shaft can be considered to be the same as that of a single shaft engine, that of the LP shaft will be quite different.

Work conducted at NGTE by Mullins (1951) studied the dynamics of the combustion process itself. This was discussed by Bauerfeind (1968) and later by Thomson (1974), who presented a semi-empirical combustion transfer function, with a pure time delay.

$$G_c(s) = \frac{1}{(\tau_c s + 1)} e^{-(T_1 + T_2)s} \qquad (7-4)$$

The combustion time constant was found to be in the range of 20-80 ms for most burners, which is considerably smaller than the shaft time constants. The pure time delay was postulated as being made up of two components: T_1 being the time taken for the fuel to enter the jet stream after leaving the burners and T_2 the time taken for it to vaporise sufficiently for combustion. The first was assumed to be a constant, in the order of 5 ms, and the second was found to be pressure-dependent. The value of T_2 is around 10 ms for a typical combustion chamber, giving a total pure time delay of 15 ms. Values of this order are also quoted by Harman (1981, chapter 7).

7.4.1 Thermodynamic Models

The main limitation of the time-constant models was the small operating range over which they were valid, which meant that they were mainly employed for controller tuning, using well established frequency-domain techniques. The next major advance was made by Saravanamuttoo and Fawke (1970) who developed a nonlinear thermodynamic model for use in the analogue and digital simulation of an Olympus engine. This model was derived from the engine physics and was able to predict the change in engine dynamics due to changing operating conditions, over the complete running range of the engine. It was assumed that the performance of each component could be modelled by its steady state characteristic and that dynamic equations could be introduced to describe the transient behaviour between the components.

The digital version of this model was validated by comparing its output with data from engine tests, using slowly ramped acceleration and deceleration input signals. A reasonable agreement was obtained between model and data, making the model suitable for both control system design and response simulation at the engine development stage (Fawke *et al.*, 1972).

The Saravanamuttoo approach was used as the basis for the thermodynamic models developed at the NGTE by Cottington (1974). The development of more complex models underlined the need for quality engine data with which they could be validated and the lack of such data at that time. This was largely due to the long test times associated with single sine testing, which made them an expensive and consequently unattractive option during the development of an engine. Typical test times of between 15 and 30 minutes were required in order to measure only 15 frequency points.

This problem was addressed in a further paper by Cottington and Pease (1979), who used MLBS signals to test a number of engines, which allowed many frequency points to be measured in one test. The amplitude of the perturbation signals was restricted to produce output speed variations of no more than $\pm$ 2% of the maximum N_H, in an effort to ensure a linear response and minimise engine wear. Cottington and Pease stressed the importance of excluding the dynamics of the fuel feed system from the estimated model, by using the measured fuel flow as the input signal.

The results were compared with those obtained from single sine tests and found to be of acceptable accuracy, with test times reduced to less than five minutes. The scatter obtained in a number of the frequency response plots was attributed to nonlinearities, though no attempt was made to quantify their effects, or to assess the uncertainty of the frequency responses. It was concluded that MLBS signals represented an attractive alternative to single sines, which should facilitate more extensive engine testing in the future.

The need to address the dynamics of the fuel feed system was discussed by Moore (1970), who studied the influence of drift and nonlinearities on the estimated impulse response of an engine. It was found that the impulse response of a Pegasus vertical take-off engine, estimated using an MLBS signal and cross-correlation, contained unexpected additional peaks at lags which did not correspond with the engine dynamics. Godfrey and Moore (1974) showed that these effects could be caused by a direction-dependent nonlinearity in the fuel feed, which means that the system dynamics depend on whether the input signal is moving in a positive or a negative direction. This was confirmed by testing the fuel feed system of an Olympus 593 independently of the engine, where a similar pattern was found.

The underlying theory to explain these effects had been previously developed by Godfrey and Briggs (1972), who derived an analytical expression for the output of a direction-dependent first-order system subject to a binary input. This led to a discrete Volterra-type expansion, with both even and odd terms expressed as weighted sums of multiples of past inputs. The convergence of the series was shown to depend on the relative value of the dynamics in each direction. If the dynamics are quite similar then only the constant, linear and second-order terms need to be taken into account; if they are very different then higher-order terms must also be included. The use of an inverse-repeat MLBS was shown to eliminate the peaks associated with the even-order terms.

A fuller discussion of the work carried out at the NGTE was presented by Onions and Foss (1982) who discussed the results obtained on a Spey Mk 502 engine. The first step in their investigation was to compare the steady state outputs of the model and engine. Significant discrepancies were found and an iterative procedure used to obtain a consistent steady state match between the output of the engine and the thermodynamic model. Dynamic response data were then gathered using MLBS signals applied at 92% N_H and 74% N_H, which represent high power and flight idle engine conditions. The results obtained at 92% N_H can be seen in Figure 7-6, which shows considerable mismatch,

particularly in the phase. A number of approaches were then attempted to improve the model fit, which included: "varying the control volumes, shaft inertias and applying fuel lags and delays". None of these was found to significantly improve the fit.

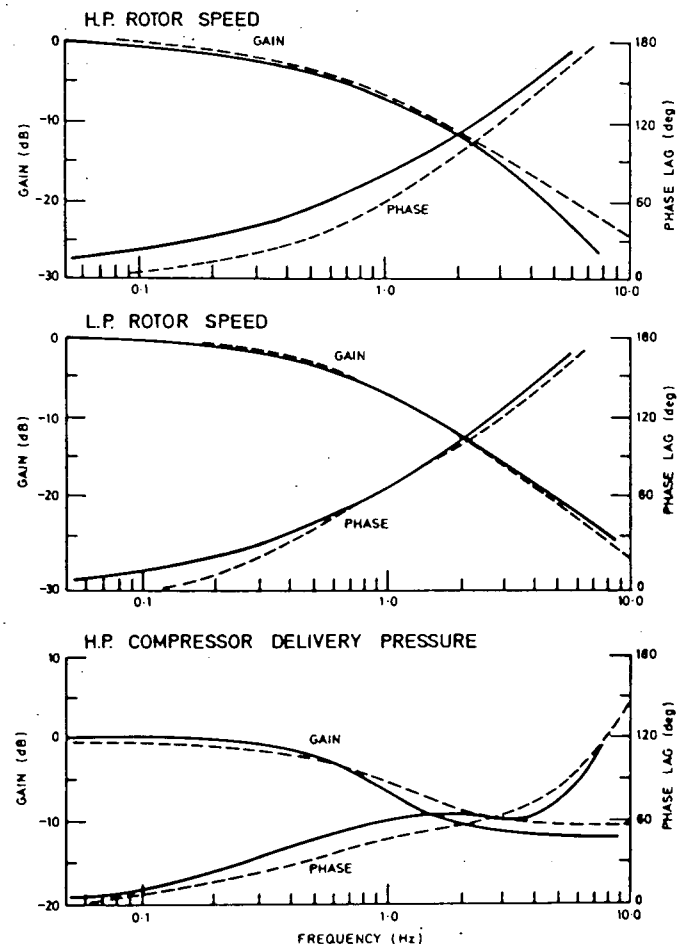

Figure 7-6. Measured and simulated frequency response functions, at 92% N_H, measured (solid) and simulated (dashed). (Reproduced from Onions and Foss (1982), with permission.)

In work conducted for Rolls Royce plc, Jackson (1988) showed that, for a given stationary operating point, the higher order nonlinear thermodynamic models derived from the engine physics could be reduced to linear models with the same order as the number of engine shafts. The models were first linearised using small perturbations, to arrive at 15 state models and then a model reduction procedure was employed. The model reduction

is possible because the thermodynamic processes in the gas stream, associated with rates of change of pressure, have time constants of less than 20 ms, while the shaft time constants lie in the range of 0.3 - 4 s.

The shaft speeds were selected as the states, which is convenient since they can be directly measured. A complete library of such models has been generated for the Spey engine across a range of operating points, with different settings of the inlet guide vanes and nozzle area (Staff at Section APD5, 1993). These linearised, reduced-order thermodynamic models will be referred to simply as the *thermodynamic models* in future discussions.

Evaluating the transfer function matrices of these state-space models allows the HP and LP shaft dynamics to be expressed in transfer function form, with common poles but different zeros. These poles and zeros are plotted for a range of operating points in Figure 7-7. It can be seen that the HP shaft has a cancelling pole-zero pair for much of the operating range, while the LP shaft zero is more distinct. This result matches that obtained by Fitchie *et al.* (1959) on the Olympus 101 and suggests that the HP shaft is effectively first-order and the LP shaft second-order. However, the second-order dynamics of the Spey LP shaft are more significant than those of the Olympus, as shown by the greater separation of the second pole and the zero. This is also true for the HP shaft at higher operating points.

The HP and LP transfer functions at an operating point of 75% N_H are shown in equations (7-5) and (7-6) and, once again, expressing them in partial fraction form illustrates the relative contribution of each of the poles to the shaft responses.

$$H_{HP}(s) = \frac{(0.318s+1)}{(1.435s+1)(0.316s+1)} = \frac{0.696}{(s+0.697)} + \frac{0.0057}{(s+3.165)} \qquad (7-5)$$

$$H_{LP}(s) = \frac{(0.364s+1)}{(1.435s+1)(0.316s+1)} = \frac{0.667}{(s+0.697)} + \frac{0.135}{(s+3.165)} \qquad (7-6)$$

The development of increased computing power, along with advances in system identification techniques, have now made it possible to apply a more systematic approach to fitting models to engine data. Recent work by Hill (1994) examined the application of

a range of time-domain approaches to estimating discrete engine models, with the aim of reducing engine test times while improving the accuracy of the estimated models and quantifying the model uncertainty.

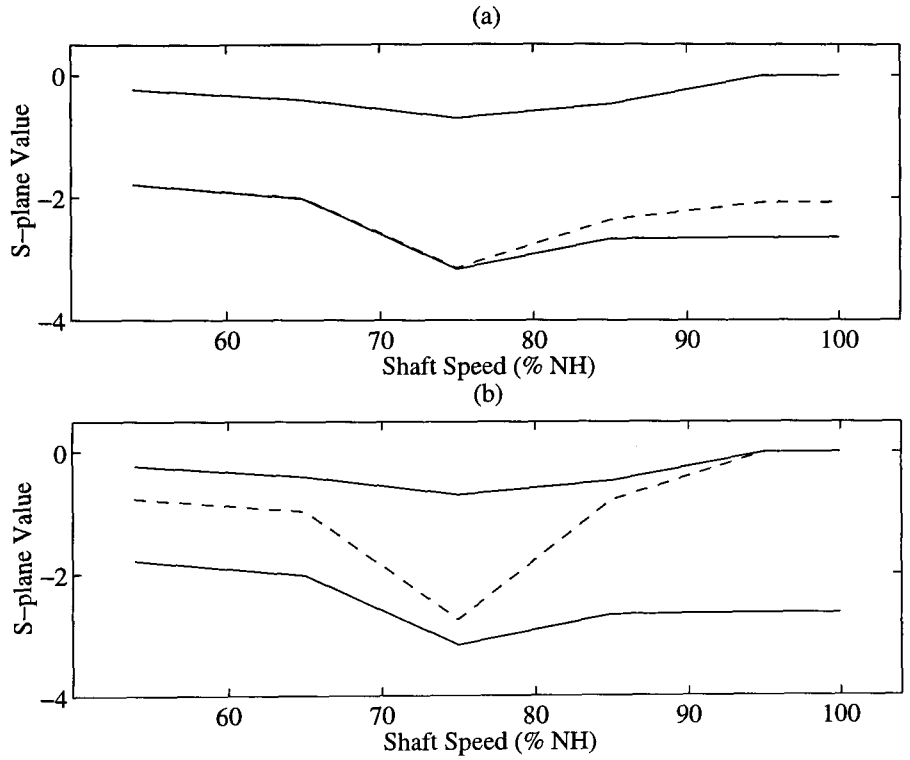

Figure 7-7. Poles and zeros of the Spey thermodynamic models (a) HP shaft and (b) LP shaft. Showing poles (solid) and zeros (dashed).

Hill examined the structure of the models derived by Fitchie *et al.* (1959) and Jackson (1988) and identified the following features: the LP and HP shafts have a common dominant time constant; the second-order effects in the models are associated with the interaction between the shafts; and the HP shaft has a near-cancelling pole-zero pair. Particular importance was placed on the identification of the second-order effects, since they had not been adequately dealt with in past work. A more detailed knowledge of these effects is now required in order to develop multivariable control strategies and also to model the shaft speed responses to *power offtakes* on the HP and LP shafts. These power offtakes are used to drive pumps and generators within the engine and the aircraft. This is currently an area of considerable interest to engine manufacturers and will be an important feature of future engine designs (Rolt, 1993).

A great deal of attention was paid to the input signal design and two contrasting approaches were examined, these being the use of wide-band MLBS signals and

optimised multisines containing only a few harmonics. The multisines were optimised using a technique which minimised the dispersion function, using the known thermodynamic models in the optimisation routine. Signals with a maximum of four harmonics were designed, which were not amenable to CF minimisation due to the signal sparsity and the uneven spacing between the harmonics.

Data were gathered on the Spey engine at an operating point of 75% N_H, using single sines, MLBS signals and multisines, with a range of input amplitudes up to a maximum of ± 20% W_f. The value of the engine pure time delay was obtained by estimating a range of first-order models with different delays and selecting the one with the smallest model error. It was concluded that the delay was one sample period, of 35 ms, or less. Discrete transfer function models were then estimated using extended least squares and modified ellipsoid bound techniques and state space models estimated using extended least squares.

The estimated second-order transfer function models each had single real negative poles, which could not be transformed into the s-domain using impulse invariant techniques, precluding their comparison with the existing thermodynamic models. This led Hill to conclude that these second-order modes were spurious and that they should be discounted. It will be shown in Chapter 9 that this effect can be caused by an incorrectly specified pure time delay or the use of a measured fuel flow in the estimation, in order to exclude the actuator dynamics from the model. The actuator band-limits the input signal before measurement, which violates the ZOH assumption for discrete models. The problems encountered with the time-domain models were discussed in a recent paper by Hill (1997).

The state-space models gave better results, in that both poles of the second-order models were positive and in reasonable agreement with the thermodynamic models. In comparing the models estimated with the MLBS signals and the optimised multisines of few harmonics, Hill concluded that the benefits obtained with the optimised multisines were too small to justify the effort involved in their design.

7.4.2 Thermal Effects

A major problem with the models previously described is the assumption that the engine components can be modelled by stationary characteristics, which is not valid for large accelerations and decelerations, termed *slam* manoeuvres. Under such conditions, *heat*

soakage effects become significant, which relate to heat transfer between engine parts, changes in blade and seal clearances and changes in combustion efficiencies and component characteristics.

One of the earliest attempts to address this problem was made by Bauerfeind (1968), who developed an engine model which incorporated the effects of heat exchange between the gas stream and the engine components. Bauerfeind also modelled the combustion dynamics and the *packing lags*, which are caused by the need for the mass contained within a given engine volume to change whenever the temperature and pressure are altered.

The effect of heat transfer between the gas stream and the compressors and turbines was incorporated into a general model of a twin spool turbofan developed by Fawke and Saravanamuttoo (1973). The model output was compared with engine data for a series of slam manoeuvres, including a *hot reslam,* where an engine running at maximum rating is decelerated and then immediately reaccelerated. Inclusion of the thermal modelling was shown to slightly improve the match between the model and the engine data.

Thomson (1974) noted that engine models which did not include thermal effects tended to under-predict engine response times to slam manoeuvres by as much as 20-30%. The Thomson model included the effect of heat absorption in the combustion chamber and the variation of compressor and turbine tip clearances with temperature. The influence of changes in compressor characteristics during transient manoeuvres was studied by Maccallum (1979, 1981), who showed that it had a greater effect in twin-shaft engines.

Tests conducted at the NGTE revealed that, for large acceleration manoeuvres, the thermodynamic model reached steady state some five to ten minutes before the engine itself (Onions and Foss, 1982). This is illustrated in Figure 7-8, which shows the response of an engine to a large step change, followed by a ramped increase in the fuel flow. The engine accelerates slower than the simulation and appears to reach a different steady state condition. In fact, the engine has not reached steady state and its response continues to rise slowly, only matching that of the model after several minutes.

Two ways of improving the large transient behaviour were then considered, which involved adding an acceleration dependent factor to the existing thermodynamic models.

The simplest method was to change the moment of inertia of the shafts and while this produced a significant improvement in the results it was subsequently discounted, as there is no physical justification for this approach.

The second method involved decreasing the combustion chamber efficiency during a large scale acceleration. This can be related to a real physical effect, that of a fall in efficiency due to over-fuelling, which is thought to be caused by the reduced atomisation of the fuel or a distortion in the relation between the reacting substances, termed the *stoichiometry*, of the combustion zone. The improvement achieved for the slam acceleration of a three shaft engine can be seen in Figure 7-9. The resulting model was validated using data gathered by introducing a fuel *spike* into the engine, which involved a rapid increase in fuel flow, followed by a constant value for around one second and then a rapid decrease.

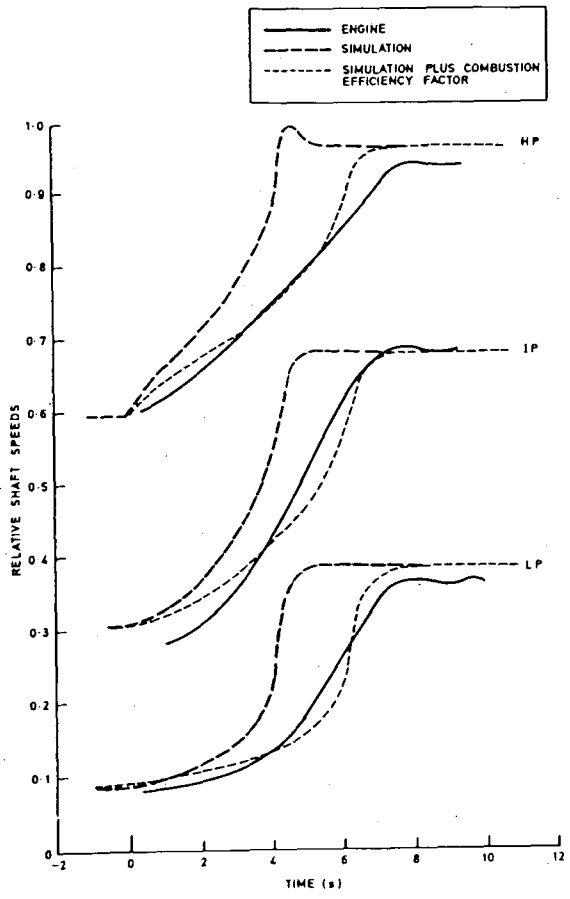

Figure 7-9. Slam acceleration at sea level static conditions.
(Reproduced from Onions and Foss (1982), with permission.)

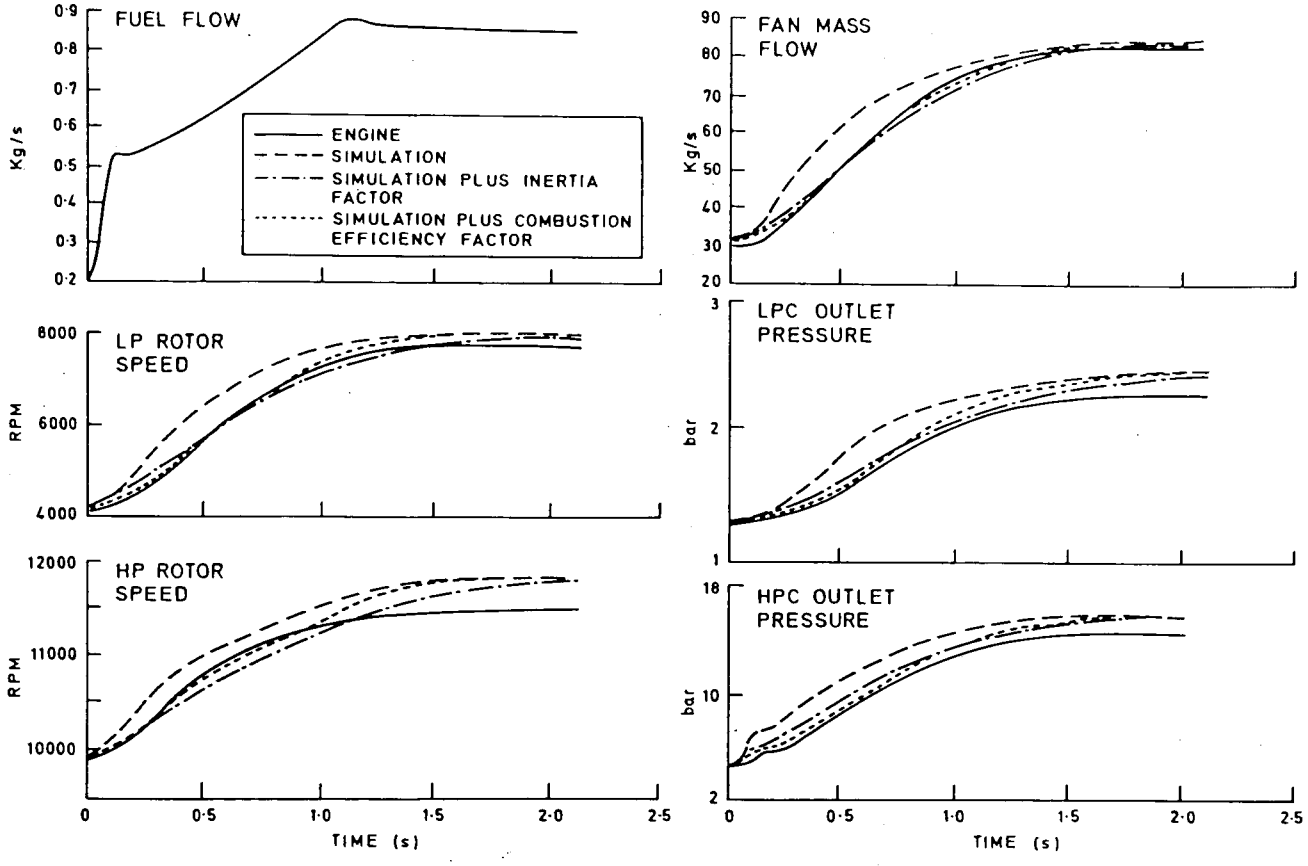

Figure 7-8. Engine response to a large transient manoeuvre.
(Reproduced from Onions and Foss (1982), with permission.)

Other approaches discussed by Onions and Foss involved a more detailed modelling of the compressor characteristics, in order to simulate and predict the onset of surge. Each approach involved a considerable increase in model complexity, resulting in two-shaft engine models with as many as 17 additional modes, when linearised. The problem of fitting such high-order models to the engine data was noted, particularly given the lack of sufficient data at that time.

A great deal of attention has been paid to the modelling and control of the compressor and turbine blade tip clearances. The variation of the clearances during a transient manoeuvre is caused by the differential expansion of the engine casing and the rotor blades, which must be kept to a minimum to maintain maximum engine efficiency. This work is of

particular interest since it provides estimates of the time constants of various thermal effects. Rotor blade expansion time constants of 5 seconds and rotor disc expansion time constants of 5 to 10 minutes are quoted as typical by Hennecke and Trappmann (1982).

A model for predicting the tip clearances of the HP compressor of a typical twin-shaft turbofan was developed by Pilidis and Maccallum (1982). The compressor model had 12 stages and the calculated time constants of various components of the fifth stage are listed in Table 7-1, at three instants in a transient manoeuvre. A later paper by Pilidis and Maccallum (1986) presented a useful summary of the different heat transfer effects encountered in a gas turbine and their relative influence on the transient performance. This work suggested that the thermal effects with the most significant influence on the shaft responses were the changes in the HP turbine seal clearances and in the compressor characteristics.

TABLE 7-1

TIME CONSTANTS DURING ACCELERATION
(from Pilidis and Maccallum (1982), with permission)

Component	Time point during transient		
	2s	6s	10s
Disc hub	108s	66s	37s
Disc outer section	12.7s	9.0s	5.5s
Blade	2.3s	1.6s	1.0s
Casing	10.4s	7.6s	4.8s

A further work by Maccallum and Qi (1989) included a discussion of the relative influence of the combustion dynamics on the response of the shaft speeds to a large scale fuel transient. A reference was made to unpublished work by Saravanamuttoo which suggested that the dynamics depend on the type of burner used to inject fuel into the engine. They concluded that the combustion dynamics could be disregarded when studying large-scale transient manoeuvres, since the combustion time constant was too fast to play a significant role and the pure time delay was also very small.

Recent work by Hermsmeyer (1996) applied a very different approach to the problem, which involved estimating an overall black-box model of the engine, rather than

physically modelling each of the different thermal effects. The model consisted of two parts, the first being a static ramp model linking the shaft speeds to the fuel feed and casing temperature across the complete engine operating range, and the second a small-signal model of the shaft dynamics, in state-affine form.

The signals used to excite the gas turbine were similar to those used by Hill and consisted of broad-band MLBS signals and multisines containing only a small number of harmonics, which were superimposed on ramped accelerations and decelerations during testing. Hermsmeyer also suggested that it might be appropriate to employ Volterra kernels in modelling the nonlinear engine dynamics, the measurement of which was discussed in Chapter 6 of this thesis.

7.5 Conclusions

A great deal of work has been conducted on the dynamic modelling of gas turbines, an overview of which has been presented in this chapter. Early work by Fitchie *et al.* (1959) on turbojet engines indicated that the fuel feed to shaft speed dynamics of a twin shaft engine can be considered as first-order for the HP shaft and weakly second-order for the LP shaft. Later work by Jackson (1980) on linearised thermodynamic models of a Spey engine showed the same pattern but also indicated that the second-order LP shaft dynamics are more significant for this turbofan design. These models have yet to be fully verified using experimental results.

It is clear that the systematic application of modern system identification techniques to this problem is still at an early stage. The work conducted to date has concentrated on the time-domain estimation of discrete models. Such models are extremely useful for purposes of engine simulation and control system design but they cannot be interpreted in terms of physical parameters if they are estimated using a band-limited input signal, due to the violation of the ZOH assumption (see Chapter 3).

One response to this problem is to significantly increase the sampling frequency, so that the ZOH assumption is acceptable, even if not precisely true. Alternatively, continuous-time models can be directly estimated in the frequency domain, in which case it is important to ensure that the BL assumption is not violated. This frequency-domain approach will be employed in the following chapters.

The application of multifrequency test signals has been motivated by the need to reduce engine test times, in order to make systematic engine testing a cost-effective option. The multifrequency signals applied in previous work were MLBS sequences and multisines made up of only a small number of harmonics. No attempt has yet been made to systematically assess the influence of the gas turbine nonlinearities on the test signals and consequently on the estimated linear models.

It is therefore clear that further work is required in this area, with the aim of:

- Applying wide-band, low CF, multisine signals to the testing of gas turbines.
- Estimating many points on the engine FRF in one test, along with a measure of the uncertainty on the estimates.
- Directly estimating *s*-domain models using frequency-domain techniques, which can be used to verify the existing *s*-domain thermodynamic models.
- Systematically assessing the influence of system nonlinearities across the selected input amplitude range and their influence on the estimated linear models.
- Estimating the engine pure time delay as part of the *s*-domain model.

The following chapters will present the results of work conducted to address each of these points. The application of a range of multisine signals will be discussed, which were designed according to the procedures outlined in the preceding chapters of this thesis. The overall aim was to verify the existing thermodynamic models of the Spey engine, with particular emphasis on identifying the second-order effects, which are thought to model the shaft interactions. The possible influence of thermal effects, which are not incorporated into the thermodynamic models, will also be investigated.

Chapter VIII

Gas Turbine Testing

Abstract — *The engine test facility is described and the problems associated with testing are discussed. Data previously gathered on the same engine are used as a priori information to aid in test signal design. A range of broad-band multisine test signals is designed and the rationale for their application is outlined. The engine tests are described and a detailed analysis is made of the gathered data, with regard to the synchronisation of sampling, process drift, noise and nonlinearities. The nonparametric frequency response functions of the HP and LP shafts are then estimated for each of the tests, along with confidence bounds, and the results compared.*

8.1 Introduction

The engine tests were conducted at the Glen sea-level test facility of the Defence Evaluation and Research Agency at Pyestock. A Rolls Royce Spey Mk 202 engine was tested, which is a typical military two-shaft turbofan, with a low by-pass ratio and a variable reheat nozzle. The engine is controlled by varying the rate of fuel flow, the angle of the inlet guide vanes, the reheat nozzle area, the reheat fuel flow and the LP compressor bleed valve position. The reheat system was inoperative during the tests and the compressor bleed valve was closed. The angle of the inlet guide vanes and the reheat nozzle area were fixed to appropriate values for the duration of the tests.

The engine speed control was operated in open loop and a perturbed fuel demand signal fed to the fuel feed system, which regulates the fuel flow to the engine by means of a stepper valve. There are a number of problems associated with the measurement of the fuel flow and shaft speeds, which will be discussed in turn. Some of these are inherent to gas turbines, others are features of the test facility, which is represented diagrammatically in Figure 8-1.

Fuel Feed System. The fuel feed system is composed of a valve driven by a stepper motor, which exhibit both linear and nonlinear dynamics and affect the actual fuel flow applied to the gas turbine. It is important to eliminate these effects from the estimated engine model and this is achieved by measuring the actual fuel flow downstream of the fuel feed, using a turbine flow meter. The nonlinear behaviour of the fuel feed systems fitted to the Pegasus and Olympus 593 engines was studied by Godfrey and Moore (1974), who proposed the use of an inverse repeat binary sequence to reduce the influence of the nonlinearity on the estimated engine model. The degree of nonlinearity present in the fuel feed system fitted to the Spey will be addressed in the current work.

Speed Measurement. The speed of the low pressure shaft is measured by counting the rotations of the turbine blades and the speed of the high pressure shaft by measuring the rotation of a gear linked to the shaft itself. A series of pulses is generated, the frequency of which depends on the shaft speed. The number of pulses in a given interval is counted and the resulting speed provided as the transducer output. The engine control computer then polls this output at each sampling instant. It should be clear that it is not possible to incorporate any AA filters into this set-up, since the conversion is from a

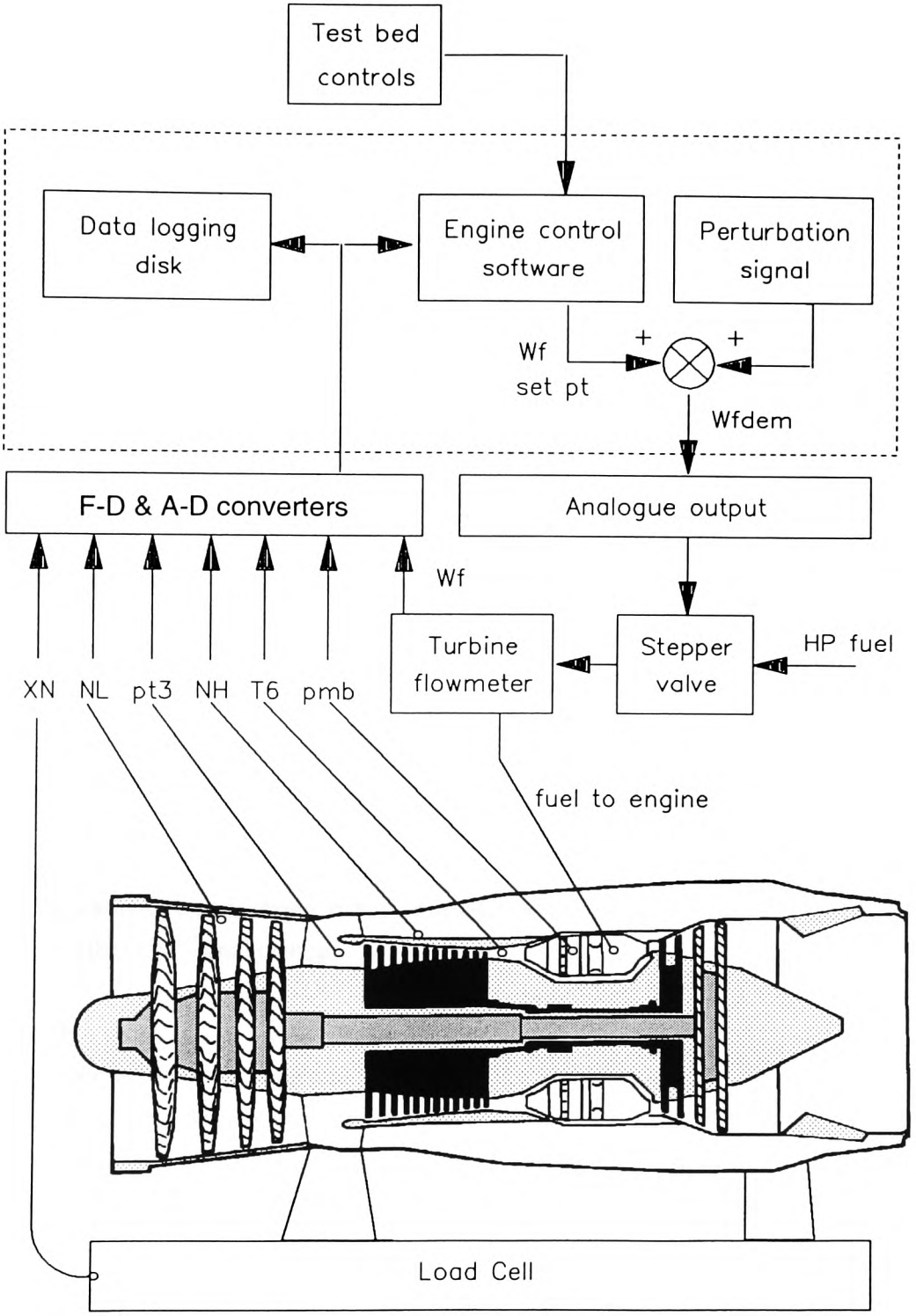

Figure 8-1. Engine test facility.
(Reproduced from Hill (1994), with permission.)

pulse frequency to a digital value (F-D) and no true analogue signal exists before digitising. This also applies to the turbine flow meter on the input.

Noise. The dominant noise sources are the turbulence in the fuel flow on the input and the vibration of the turbine blades and jitter in the gears on the outputs. This means that the input noise cannot be considered as simply measurement noise, since it also passes through the system and influences the outputs.

Engine Control Computer. The number of samples which can be logged in one test is limited to 15,000 and for this reason only the demanded fuel flow, the measured fuel flow and the two shaft speeds were recorded. Recording the demanded fuel flow allows the generation of the test signals to be checked and also allows the fuel feed dynamics to be studied. The *digital-to-analogue converter* (DAC) at the computer output and the samplers at its input can be clocked at different rates, allowing signal reconstruction and data sampling at different rates.

8.2 Models Based on Previous Data

The data gathered on the same engine by Hill (1994) were available for analysis before testing commenced and proved an invaluable aid in designing appropriate test signals. The data are summarised in Table 8-1 and consist of five single sines along with a number of multisine and MLBS tests. Problems with the data mean that the models estimated in this section cannot be considered as fully reliable and they will only be used as a general guide for further test signal design.

The ability to combine data from different tests is an attractive feature of frequency-domain identification and this allowed the single sine tests to be combined into one global data set. An attempt was also made to incorporate the frequency data from each of the multisine tests but it was found that the engine FRF estimated with these data was quite different from that estimated with the single sines. The variation of the FRF points would have made it difficult to fit a model to the combined data and the multisine data were omitted for this reason.

A problem was encountered, in that the signals had been sampled at frequencies which were not integer multiples of their fundamentals, thus introducing leakage errors into the FFT. The leakage was minimised by calculating the FFT across data lengths which were as close as possible to the signal periods but systematic errors nonetheless remained.

Table 8-2 shows the variation of the cost function, pure time delay and poles and zeros, for HP shaft models of increasing order, estimated using the combined sinusoid data set. The expected value of the cost function K_{min} is calculated from equation (2-39).

TABLE 8-1

SUMMARY OF HILL ENGINE TEST DATA

Signals	Frequencies (Hz)
Single sine tests	0.027, 0.1, 0.357, 0.6, 2.0
Multisines	(1) 0.15, 0.75, 14.3 (2) 0.04, 1.88, 2.44, 4.74 (3) 0.20, 2.71, 14.3 (4) 0.37, 1.83, 4.74
MLBS at different amplitudes	$f_{min} = 0.028$; $\Delta f = 0.056$; $f_{-3dB} = 12.656$

TABLE 8-2

HP SHAFT MODELS ESTIMATED USING COMBINED SINUSOID DATA

Order	Cost Function (K)	K_{min}	Delay (ms)	Zeros	Poles
0/1	92	3.5	10	—	-0.7141
1/1	57	3	85	-12.55	-0.6971
1/2	21.3	2.5	13	-0.5327	-0.4228 -0.8715
2/2	21.1	2	13	-0.5327 2.660e+6	-0.4228 -0.8715

The cost function is considerably larger than the expected value for all the model orders, indicating the presence of systematic errors in the data. It can be seen that the major drop in the cost function occurs with the one zero, two pole (1/2) model, which has a credible

value of pure time delay, a dominant pole at s = -0.87 and a close pole-zero pair at lower frequency. Increasing the model order to 2/2 has little effect on the cost function and introduces a zero at a frequency considerably outside the test bandwidth. A similar effect occurred if further poles were added to the model.

The standard deviations of the poles and zeros are not quoted in Table 8-2, since no meaningful estimates of the noise variances could be made with these data. The input and output noise variances were thus set to a nominal value for each input and output frequency, proportional to the noise power calculated from data gathered by Hill while no additional perturbation signal was applied to the engine, termed a *signal-off* test. This ensured that the noise weighting was broadly correct, which is important for interpretation of the estimator cost function.

Cross-validation was performed using one of the MLBS tests, the problem in this case being that the signal was not band-limited before sampling, which meant that aliasing errors were introduced. The first 25 frequencies of the signal were used in the estimation and the results are shown in Table 8-3, where the influence of the aliasing error can be seen in the large size of the cost function, compared to its expected value. The value of delay had to be fixed to 13 ms in the 1/2 and 2/3 models, in order to obtain convergence of the estimator. Despite these problems, the estimator is picking out a common mode to that found with the combined sinusoid data, at around s = -0.8.

TABLE 8-3

HP SHAFT MODELS ESTIMATED USING MLBS DATA

Order	Cost Function (K)	K_{min}	Delay (ms)	Zeros	Poles
0/1	1.5e+7	23.5	12	—	-0.7162
1/1	1.3e+7	23	54	-22.72	-0.7022
1/2	1.2e+7	23	13 (fixed)	-0.1393	-0.0966 -0.7813
2/2	1.0e+7	22.5	13 (fixed)	-0.1339 -3.386e+3	-0.0918 -0.7793

It is clear that these results must be treated with considerable caution, due to the systematic errors introduced by leakage and aliasing. Comparing the model cost functions with their expected values suggests that the models obtained with the combined sinusoid data are more reliable and should be used as the basis for signal design. Models of the LP shaft were also estimated with these data and are shown in Table 8-4. Once again, the 1/2 model is an attractive candidate for selection.

The 1/2 models estimated on the combined sinusoid data have poles at 0.07 Hz and 0.14 Hz for the HP shaft and at 0.08 Hz and 0.27 Hz for the LP shaft. If these values are compared with the signal bandwidths listed in Table 8-1, it appears that most of the single sine and multifrequency signals employed in the Hill tests were designed with power at frequencies much higher than the system bandwidth. Such a wide test bandwidth was required in those initial tests, in order to ensure that no dynamic effects were missed. It is now possible to focus more specifically on the frequencies where dynamics have been detected.

TABLE 8-4

LP SHAFT MODELS ESTIMATED USING COMBINED SINUSOID DATA

Order	Cost Function (K)	K_{min}	Delay (ms)	Zeros	Poles
0/1	701	3.5	-9	—	-0.9431
1/1	685	3.5	11 (fixed)	-46.97	-0.9382
1/2	17	2.5	11	-0.9487	-0.5361 -1.675
2/2	15	2	57	-0.8642 -19.32	-0.5155 -1.545

The design of optimal signals was discussed in Chapter 3, where it was suggested that they can be used to indicate the general regions where the test signal power should be placed. To this end, the 1/2 models estimated using the combined sinusoid data were used as a basis for optimal signal design, under the assumption of Gaussian noise on the input and output, using the technique proposed by Schoukens *et al.* (1993a).

The amplitude spectrum of the optimised signal for the HP shaft is shown in Figure 8-2, after 50 iterations of the algorithm. It can be seen that the maximum power is placed in regions around 0.03 Hz and 0.14 Hz. In a similar fashion, the routine placed the maximum power in regions around 0.04 Hz and 0.23 Hz for the LP shaft. The optimal input designs place the power on the initial flat portions of the frequency responses and close to the break frequencies of the dominant poles but none on the roll-off. This may appear strange but is consistent with the results presented by Schoukens *et al.* (1993a, pp. 147-149) for a bandpass system and by Van den Eijnde and Schoukens (1991) in a comparative study of two optimisation methods.

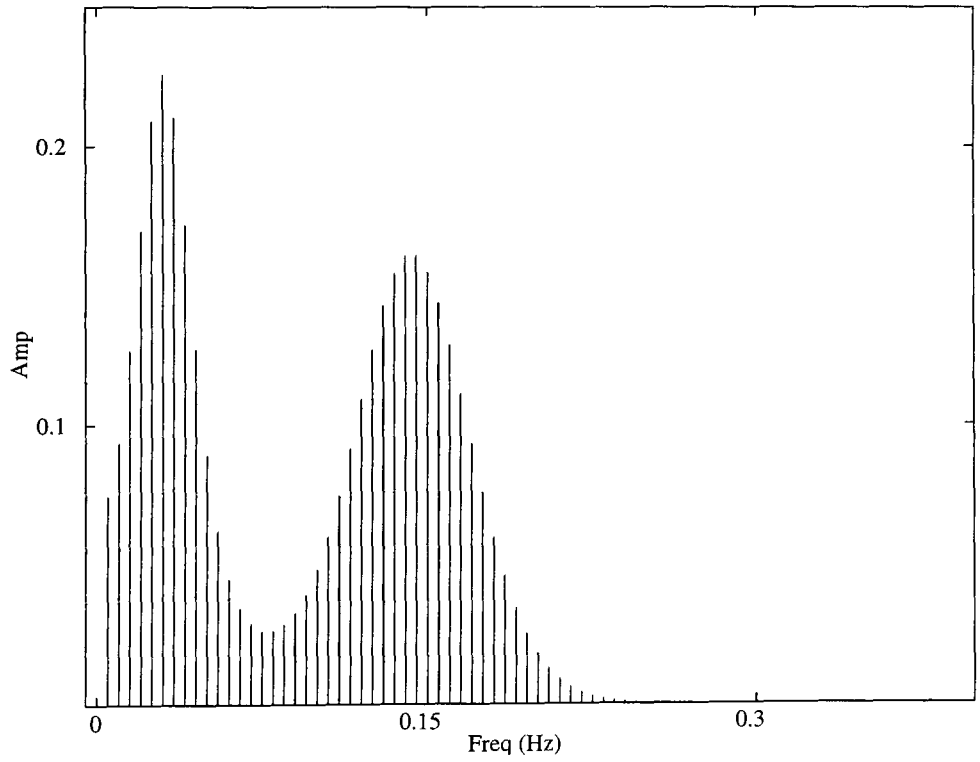

Figure 8-2. Amplitude spectrum of optimised signal for HP shaft model, after 50 iterations.

8.2.1 Fuel Feed System

The data also allowed models of the fuel feed system to be estimated. The single sine data were first examined, in order to assess the linearity of the system across the input amplitude range. Only very small amounts of output power were detected at the second and third harmonics, which suggested that the fuel feed could be considered as linear for this input range.

The single sine and multisine data were not suitable for parametric modelling, since they contained insufficient frequencies around the fuel feed system break-points. The MLBS signals were therefore employed and since these data exhibited significant aliasing and leakage errors in the frequency domain, the models were estimated in the time domain. The ZOH assumption was not violated in this case, since the input was the piece-wise constant electrical signal from the DAC and the output was the measured fuel flow.

An initial estimate of the pure time delay was obtained using cross-correlation techniques and found to be around four sample periods, or 0.14 s. Output error models were then estimated, employing the method described in Chapter 2, using MLBS data with an input amplitude of $\pm$ 10% W_f. The best model fits were found with a time delay of three samples and the variation of the cost function with model order is shown in Table 8-5, which indicates that a 1/3 structure is sufficient to model the dynamics of the fuel feed. This model was cross-validated using MLBS data with an input amplitude of $\pm$ 4% W_f and very similar results were obtained.

The poles and zeros in Table 8-5 are those of the estimated discrete models, which can be transformed into the s-domain using the impulse invariant transform. The equivalent s-domain model to the discrete 1/3 model has a real pole at 2.7 Hz and complex poles at around 6 Hz. This means that the lowest fuel feed break-point is only ten times higher than the highest engine break-point. This is a far lower ratio than that normally encountered between the system dynamics and those of the actuators and sensors. It reinforces the need to exclude the fuel feed dynamics from the estimated model by using the measured fuel flow as an input signal.

TABLE 8-5

DISCRETE FUEL FEED MODELS

Order	Cost Function (V)	Zeros	Poles
0/1	54	—	0.7333
0/2	22	—	$0.5808 \pm j0.3484$
1/2	4.23	-1.8623	$0.3705 \pm j0.3064$
1/3	2.42	-1.4921	0.5525 $0.2050 \pm j0.5297$
2/3	2.39	-1.5397 0.1010	0.5754 $0.2302 \pm j0.5297$

8.2.2 Summary

Problems of leakage and aliasing were encountered with these data, which were previously gathered on the Spey engine. This means that the estimated models must be treated with considerable caution. They can, however, be used to provide a general guide to the main system dynamics and indicate the following:

- Frequency-domain estimation using the combined sinusoid data suggests second-order models with poles at 0.07 Hz and 0.14 Hz for the HP shaft and at 0.08 Hz and 0.27 Hz for the LP shaft.
- Optimised multisines designed using the above models have maximum power at 0.03 Hz and 0.14 Hz for the HP shaft and 0.04 Hz and 0.23 Hz for the LP shaft.
- Discrete time-domain estimation suggests that the fuel feed system is third-order, with an equivalent s-domain pole at 2.7 Hz and complex poles around 6 Hz.

8.3 Test Signal Designs

The frequency-domain estimation of s-domain models must be conducted under the BL assumption, whereby the measured signals are band-limited before sampling and then sampled at more than twice their highest frequency. The lack of AA filters in the gas turbine test set-up might thus seem to preclude the use of frequency-domain techniques and it is important to address this point before proceeding to discuss the signal designs.

This would certainly be a problem if binary signals were used, since they do not have band-limited spectra. However, it is not the case if multisine signals are used, since the total signal power can be concentrated in the frequency range of interest. The only higher frequency components in the signals will be the frequency butterflies, centred at multiples of the signal reconstruction frequency, at the output of the DAC.

In this case, it is possible to utilise the low-pass characteristics of the fuel feed system to filter out these components, if a sufficiently high reconstruction frequency is used. By using the fuel feed system as a reconstruction filter the multisines can thus be strictly band-limited before application to the engine and the only aliasing present will be due to high frequency noise.

If a common clock were used for both signal reconstruction and data sampling, increasing the reconstruction frequency would have the effect of increasing the number of samples gathered for a given signal. Since the number of data points which can be logged on the test set-up is extremely restricted, this would be highly undesirable. The problem can be overcome by clocking the signal reconstruction and data sampling at different rates, while taking care to select the data sampling period to be an integer multiple of the signal reconstruction period.

The results obtained from the Hill data can be used as *a priori* information in designing a range of wide-band multisine test signals for frequency-domain estimation. It is clear from the summary presented in Section 8.2.2 that a signal bandwidth of 0.01 - 0.6 Hz is sufficient to cover the break-points of the gas turbine poles, while at the same time being significantly lower than the break-points of the fuel feed poles.

A range of signals was designed with the above considerations taken into account, along with the need for sufficient harmonics to allow CF minimisation and the detection of

modelling errors. The designs can be grouped into three categories, in accordance with their intended applications, which are: linear modelling, model cross-validation, and the detection and modelling of nonlinearities.

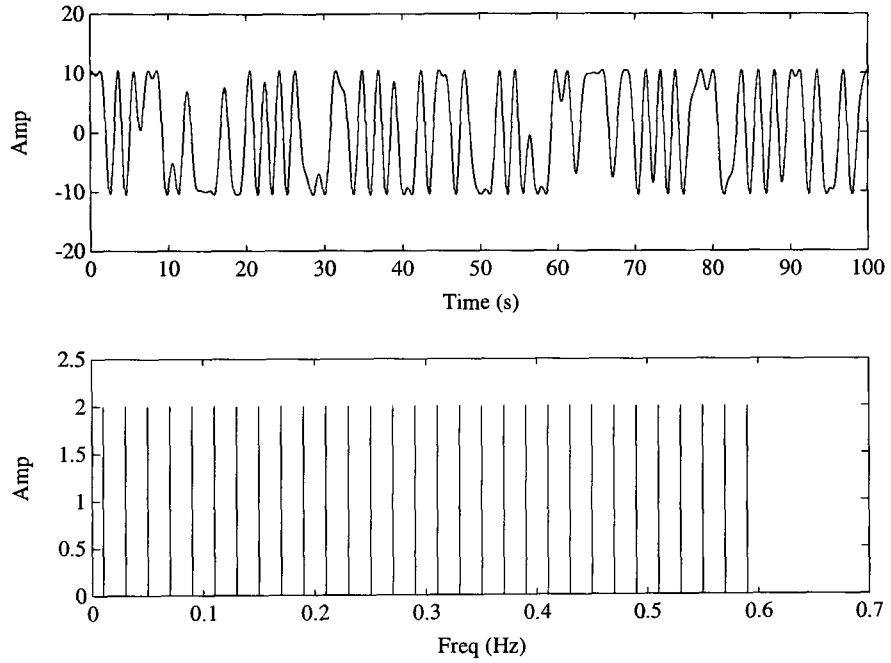

Figure 8-3. 30 odd harmonic, low CF multisine.

Linear Modelling. A 30 odd harmonic multisine with a fundamental frequency of 0.01 Hz was designed as the general signal for linear identification, the harmonic vector of which is given in equation (8-1). Such a signal is immune to even-order nonlinear effects and is amenable to CF minimisation. The application of the $L\infty$ approach proposed by Guillaume *et al.* (1991) resulted in a CF of 1.35 and the resulting signal is shown in the time and frequency domains in Figure 8-3. This signal was to be injected at amplitudes of ± 2%, ± 5% and ± 10% W_f, in order to study the influence of noise and nonlinearities on the measurements.

$$\text{Odd Multisine} \qquad \mathbf{i} = [1 \quad 3 \quad 5 \quad 7 \quad 9 \quad 11 \quad \ldots \quad 59] \tag{8-1}$$

A signal of 30 consecutive harmonics was also designed, with a fundamental frequency of 0.02 Hz, in order to have a bandwidth similar to that of the odd harmonic signal. This was included to allow the effect of any even-order nonlinearities to be studied. The time-domain amplitude of the consecutive signal was adjusted so that it had the same total

power as the ± 10% W_f odd harmonic signal, so that the SNR would be the same in both cases. Since the CF of the signals was very similar this involved only a small adjustment in the signal magnitude, to ± 10.43% W_f.

Cross-Validation. The results obtained with the 30 odd harmonic signal were to be cross-validated using a 15 odd harmonic signal with the same bandwidth but with its power concentrated at low frequencies. This will be termed a *quasi-log multisine* since it has a spectrum which is similar in appearance to that of a log-tone signal. The harmonic vector of the signal is given in equation (8-2) and the design allows more power to be injected at low frequency, for a given maximum time domain amplitude, as shown in Figure 8-4.

Quasi-log Multisine $\mathbf{i} = [1 \quad 3 \quad 5 \quad 7 \quad \ldots \quad 37 \quad 43 \quad 51 \quad 59]$ (8 – 2)

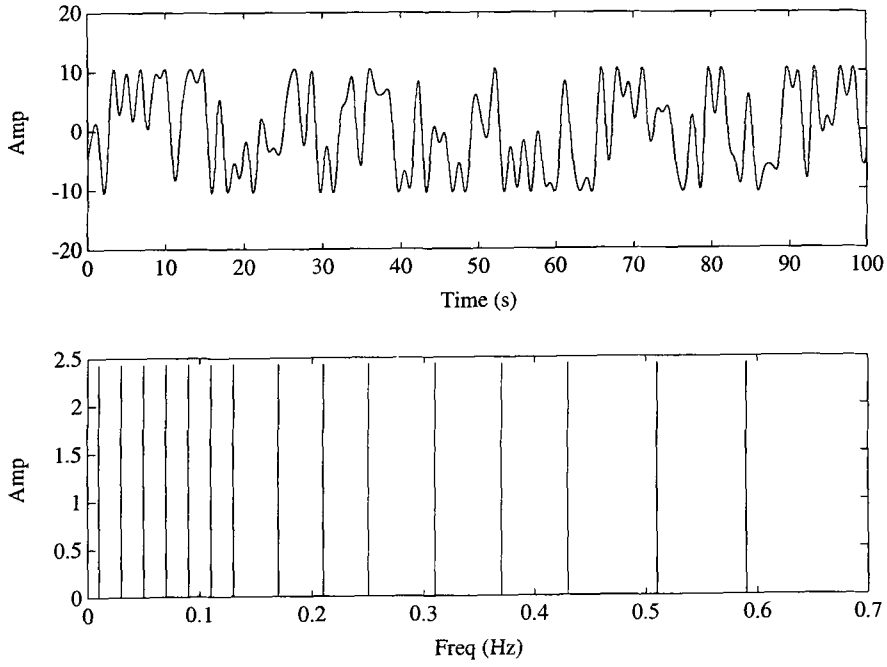

Figure 8-4. 15 quasi-log harmonic, low CF multisine.

Nonlinearities. A multisine was designed with every even harmonic excluded, along with every other odd harmonic, in order to detect nonlinear contributions between the test frequencies at the fuel feed and engine outputs. The signal is termed an odd-odd multisine

and selecting a fundamental of 0.01 Hz gives 15 harmonics within the same bandwidth as the 30 odd harmonic signal. The time and frequency plots of the signal are shown in Figure 8-5 and its harmonic vector is given in equation (8-3).

Odd-odd Multisine $\mathbf{i} = [1 \quad 5 \quad 9 \quad 13 \quad 17 \quad \ldots \quad 57]$ (8-3)

Finally, a five harmonic signal designed to measure points on a second-order Volterra kernel was included, in the hope of making an initial measurement of the HOFRF. Application of this signal was extremely speculative, since insufficient information existed prior to the tests with regard to the nature or degree of the nonlinear behaviour of the gas turbine. The signal consisted of the first five harmonics of the vector presented in Chapter 6 as equation (6-20).

Volterra Multisine $\mathbf{i} = [3 \quad 13 \quad 25 \quad 43 \quad 57]$ (8-4)

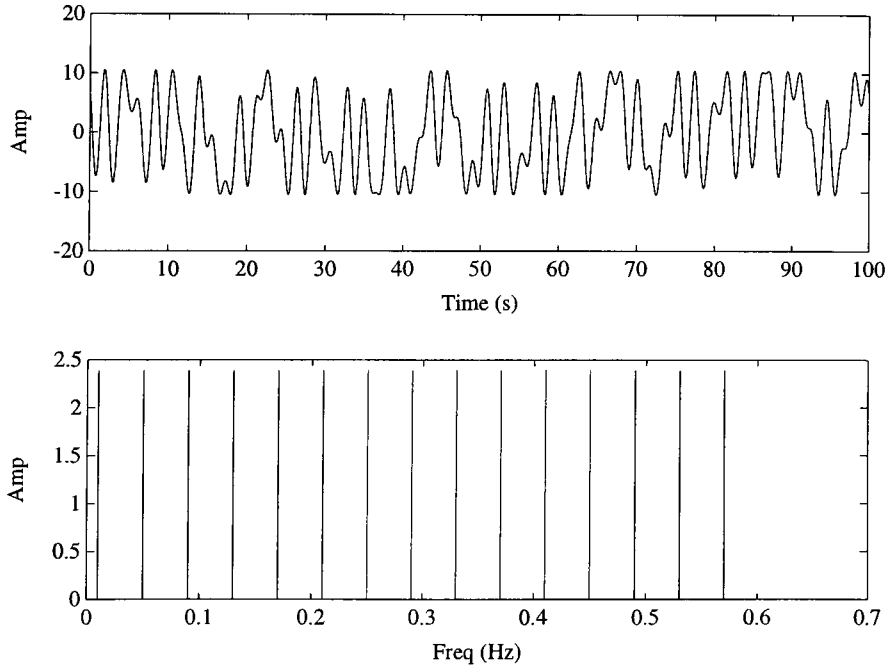

Figure 8-5. 15 odd-odd harmonic, low CF multisine.

A complete description of the tests conducted is given in Table 8-6, where the signals are listed in the order in which they were applied to the gas turbine. An incorrect value of

reconstruction frequency was used for Test 7, resulting in a slight alteration in the signal fundamental. Since the change was small and the sampling frequency was also altered in direct ratio to the reconstruction frequency, this did not present a problem.

It can be seen that a reconstruction frequency of 20 Hz was selected for most of the signals, which is seven times higher than the lowest fuel feed break-point. This should ensure that the frequency butterflies present at the output of the DAC are significantly attenuated at the output of the fuel feed. The data sampling rates are at least eight times the maximum signal frequency in all cases. A settling time of 20 s was allowed before data gathering commenced in each test. This is approximately eight times the slowest time constant of the models estimated from the previous data.

TABLE 8-6

SUMMARY OF TEST SIGNAL DESIGNS

Test	Signal Harmonics	CF	Amp. ($\pm$ % W_f)	f_{min} (Hz)	f_{max} (Hz)	f_{recon} (Hz)	f_s (Hz)	T (s)	T_{set} (s)	T_{tot} (s)	Points logged
1	30 odd	1.35	10	0.010	0.590	20	5	100	20	620	3000
2	30 odd	1.35	5	0.010	0.590	20	5	100	20	620	3000
3	30 odd	1.35	2	0.010	0.590	20	5	100	20	620	3000
4	15 quasi-log	1.57	10	0.010	0.590	20	5	100	20	620	3000
5	15 odd-odd	1.61	10	0.010	0.570	20	5	100	20	620	3000
6	5 Volterra	2.25	10	0.010	0.570	20	5	100	20	620	3000
7	30 consec	1.41	10.43	0.0222	0.666	44.444	11.111	45	20	290	3000

f_{min} - fundamental frequency

f_{max} - maximum frequency

f_{recon} - reconstruction frequency

f_s - sampling frequency

T - signal period

T_{set} - system settling time

T_{tot} - total test time ($6T + T_{set}$)

Points logged - per channel

8.4 Data Analysis

Tests were conducted at an operating point of 75% of the maximum N_H, which corresponds to a flight idle condition. The engine was stabilised at this point before switching the fuel flow control to open loop and then allowed to run for several minutes to ensure that all thermal transient effects had died away, a condition known as *heat soaked*. The signals were applied at the amplitudes listed in Table 8-6 and six periods of each were logged. Sufficient time was allowed after the application of each signal for the gas turbine to reach dynamic steady state before data gathering commenced.

Six periods were recorded in order to allow signal averaging and also to obtain meaningful estimates of the noise variances and coherence during the averaging process. A detailed analysis was made of the gathered data before proceeding with parametric estimation. All plots and discussions in the following sections refer to Test 1, with the 30 odd harmonic signal at an amplitude of ± 10% W_f, unless otherwise stated.

8.4.1 Synchronisation

It is important that the signal reconstruction and sampling clocks are synchronised when conducting measurements for frequency-domain identification, since a deviation in either clock will result in incomplete signal periods being sampled. The quality of the synchronisation can be checked by examining the circular covariance of the measured signals, defined in equation (3-22), which is plotted for the 30 odd harmonic input signal in Figure 8-6.

The periodicity of the input is clear, and the detailed plots of the first and fifth lags indicate that the synchronisation is very good. The maximum deviation between the clocks must be less than 0.025%, this being half a sampling interval (0.2/2) in four signal periods (4x500x0.2).

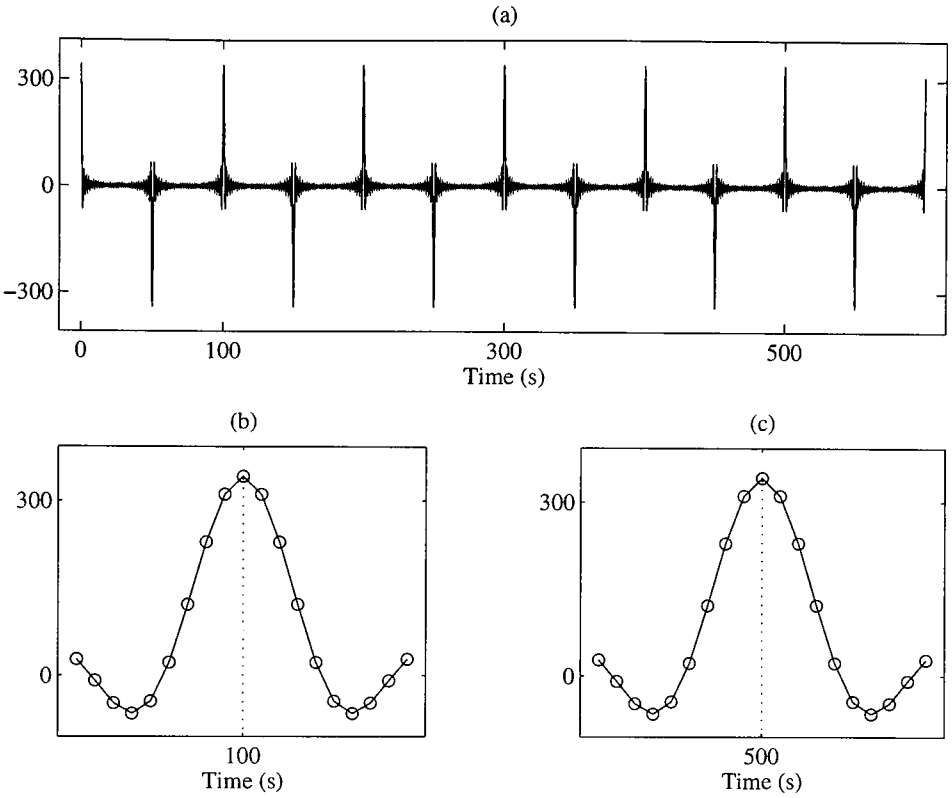

Figure 8-6. Circular covariance of input signal,
(a) across complete data length, (b) first lag and (c) fifth lag.

8.4.2 Drift and Repeatability

The possible presence of a drift in the operating point during the experiment was checked by calculating the mean values of the measured fuel flow and the shaft speeds across each test. The presence of significant drift would affect the frequency-domain estimation, since low-frequency transients would be introduced which would corrupt the Fourier coefficients of the measured data. The results are presented in Table 8-7, which shows a slight upward trend with each test. Since the increase was at most a change of 1.2% across nearly two hours of test time, it was concluded that drift effects would not significantly affect the results.

Application of periodic signals also allows the repeatability of the system response to be checked, by plotting the outputs period by period, with their mean values removed. This is shown in Figure 8-7 for the 30 odd harmonic signal, where the repeatability of the response is clear. The presence of an outlier is also revealed, towards the end of one of

the periods on the HP shaft output. A similar outlier was detected on the HP shaft output of Test 3 and both were eliminated by substituting a value calculated through linear interpolation between the previous and subsequent data points.

TABLE 8-7

MEAN VALUES OF MEASURED FUEL FLOW AND SHAFT SPEEDS

Test	Signal	$\overline{W_f}$ (cc/s)	$\overline{N_H}$ (% max N_H)	$\overline{N_L}$ (% max N_H)
1	30 odd ± 10%	252.4	74.8	40.6
2	30 odd ± 5%	252.4	75.1	40.8
3	30 odd ± 2%	252.5	75.2	40.9
4	15 quasi-log	252.9	75.3	40.9
5	15 odd-odd	252.9	75.4	41.0
6	5 Volterra	253.1	75.5	41.1
7	30 consecutive	253.1	75.5	41.1

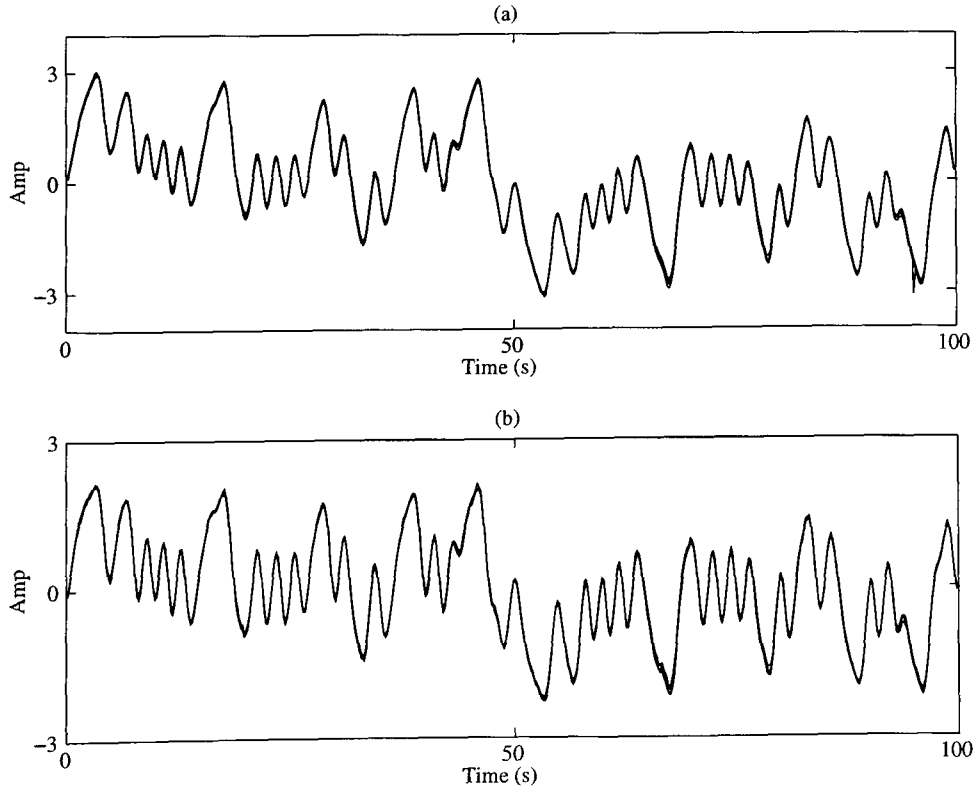

Figure 8-7. Outputs of (a) HP shaft and (b) LP shaft, plotted period by period.

8.4.3 Noise Analysis

The data were converted to the frequency domain using the FFT and the means, variances and covariance were estimated as described in Chapter 3. The variances and covariance are plotted for the 30 odd harmonic HP shaft data in Figure 8-8, which shows stronger noise variance at low frequency and also a larger covariance at those frequencies. The significant covariance is to be expected since the dominant noise source at the input is process noise, which also passes through the system and is hence correlated with the output noise. The presence of large peaks in the noise variances confirms a pattern found in previous tests.

A similar noise spectrum was obtained with the 15 quasi-log harmonic signal, which was to be used as cross-validation data, as shown in Figure 8-9. The estimated noise variances and covariance at the test frequencies will be used as *a priori* information in the parametric modelling presented in Chapter 9.

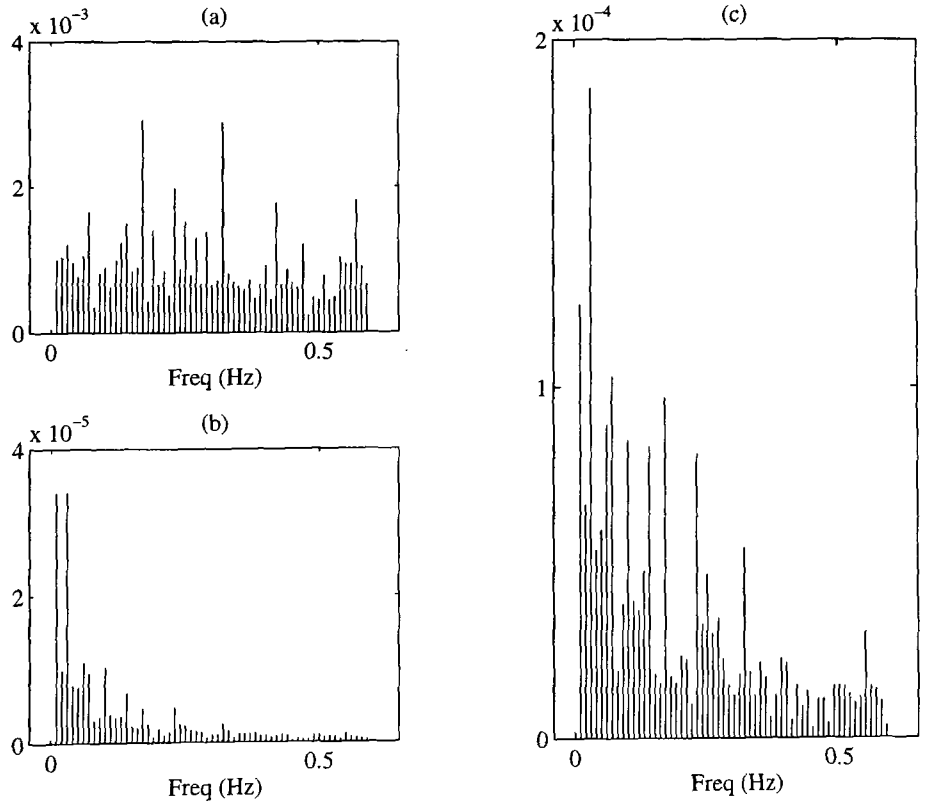

Figure 8-8. Noise variances on (a) input and (b) output, and (c) absolute value of covariance, for 30 odd harmonic signal, HP shaft.

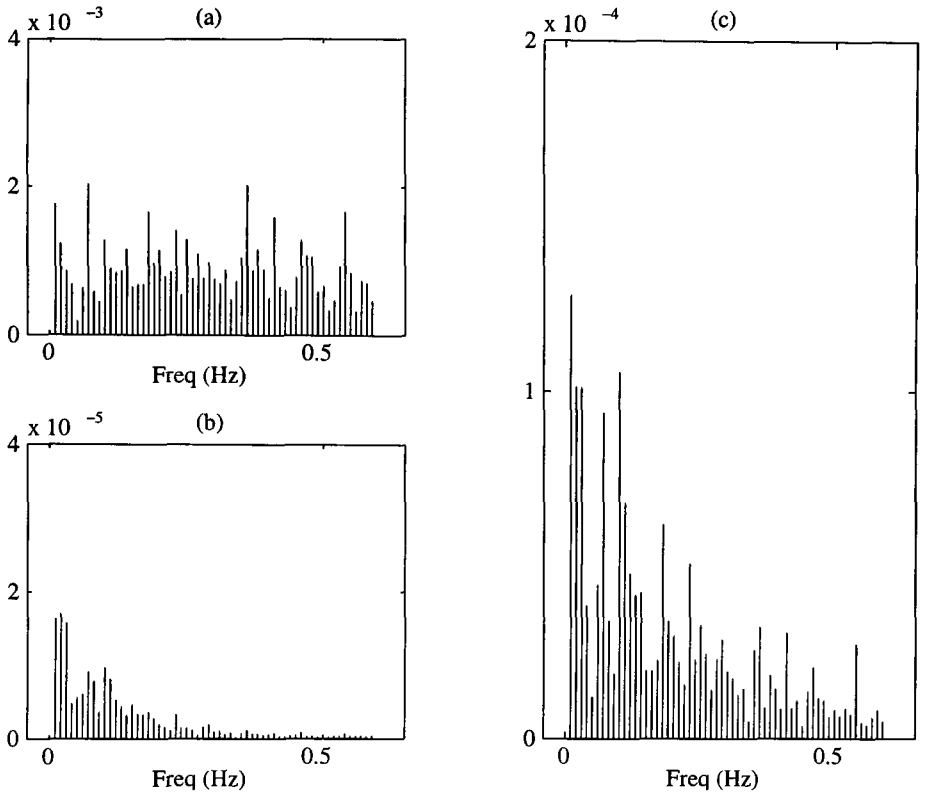

Figure 8-9. Noise variances on (a) input and (b) output, and (c) absolute value of covariance, for 15 quasi-log harmonic signal, HP shaft.

The SNRs were also calculated, beginning with an estimate of the raw time-domain SNR using equation (3-23). The improved SNRs achieved by excluding the noise lines and by averaging over six periods were then calculated using equations (3-28) and (3-29) and the results are shown in Table 8-9.

The SNRs after averaging are very good, with values of at least 40 dB in each case. According to equation (3-27) the exclusion of the noise lines should have improved the time-domain SNRs by around 9 dB, ($10.\log_{10}(500/60)$), and an improvement of this order was obtained for the input frequencies. The improvement was not as great for the outputs, since the output noise is not white and its power is concentrated in the test signal bandwidth.

TABLE 8-9

SIGNAL-TO-NOISE RATIOS

	30 Odd Harmonics			15 Quasi-log Harmonics		
SNR (dB)	Input	HP	LP	Input	HP	LP
Time domain	24	29	27	24	28	27
Excluding noise lines	33	34	34	35	37	37
Averaging	40	42	42	43	45	44

8.4.4 Nonlinearities

The presence of nonlinearities was investigated by examining the FFT of the odd-odd signal at the input and HP shaft output, shown in Figure 8-10. The input harmonics are shown in red, the excluded even harmonics in green, the excluded odd in cyan and all the other FFT lines are shown in black. There appear to be some excluded even harmonics rising above the noise floor on the output but it is difficult to conclude that these are indeed periodic nonlinear contributions.

A clearer picture is obtained from the nonlinear coherence, calculated using equation (4-15), which is plotted in Figure 8-11, along with the $1/M$ bound. It is seen that only one frequency has a significant coherence in the input spectrum while a series of even harmonics have a significant coherence in the output. The odd coherence is consistently much lower on both the input and output. This suggests that the fuel feed can be considered as linear over this range, while a weak even-order nonlinearity may be present in the gas turbine. A similar result was obtained with the LP shaft which suggests that it will be possible to eliminate the influence of nonlinearities on the estimated linear model simply by using an odd harmonic signal.

This conclusion was checked by applying the compensation method derived in Chapter 5 to the 30 odd harmonic multisine data gathered at amplitudes of $\pm$ 10% W_f and $\pm$ 5% W_f. The influence of any third-order nonlinearity was quantified by calculating the complex error E between the compensated and uncompensated FRFs, as defined in equation (4-2),

and was found to be negligible. It was also concluded that the engine was exhibiting insufficient nonlinear behaviour across the input amplitude range of the test signals to allow nonlinear modelling using the Volterra measurement signal.

Using the odd multisine data for linear estimation should therefore give high quality estimates of the engine dynamics, unaffected by nonlinearities. In practice, it was not possible to apply a pure odd signal to the gas turbine, since process noise was present at the engine input. This meant that the input signal had small amounts of power at the even harmonics, which would interact with the even nonlinearity to create distortion at the odd excitation harmonics. Since the input SNR was high and the nonlinear influence was very low it was expected that this effect would be negligible.

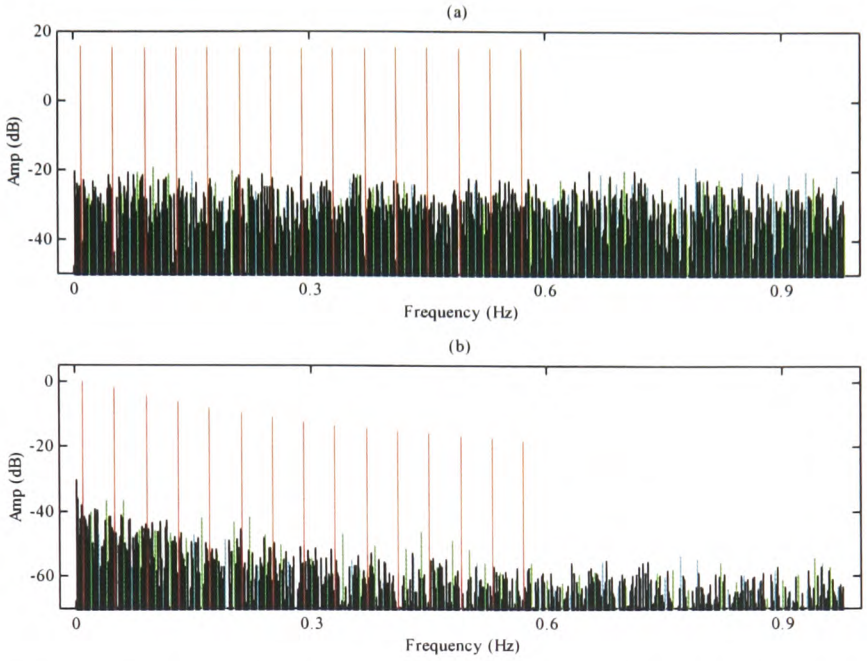

Figure 8-10. Spectrum of odd-odd multisine at (a) input and (b) HP shaft output. Input harmonics (red), excluded even harmonics (green), excluded odd (blue). All other frequencies shown black.

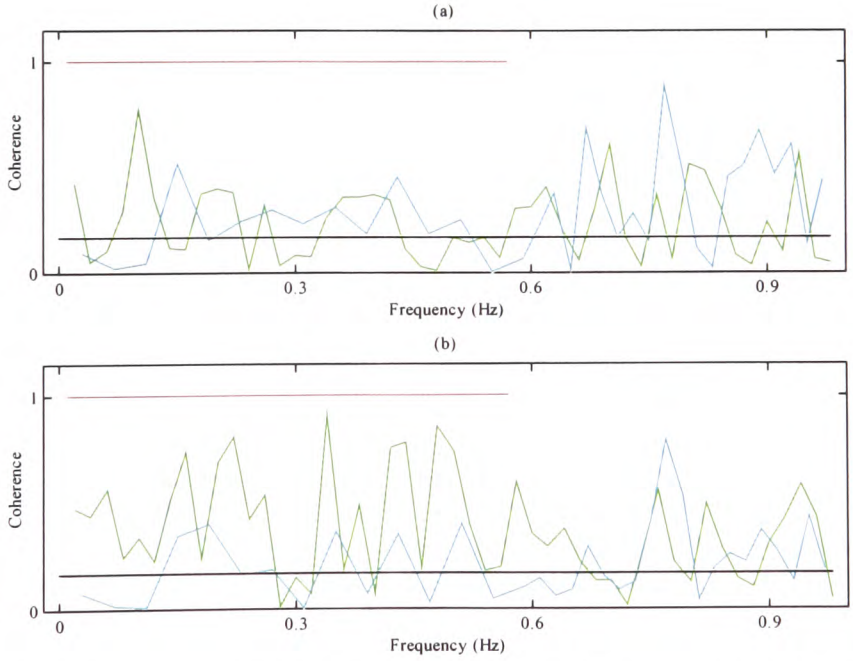

Figure 8-11. Nonlinear coherence of odd-odd multisine at (a) input and (b) HP shaft output. Input harmonics (red), excluded even harmonics (green), excluded odd (blue). With $1/M$ bound shown black.

8.4.5 Frequency Response Functions

The FRFs of the HP and LP shafts were estimated for each of the tests, using the H_{EV} estimator defined in equation (2-31), and are shown in Figures 8-12 and 8-13. It can be seen that the bandwidth of the test signals was well chosen, covering a drop of 17 dB on the HP shaft and 13 dB on the LP shaft, with the phase shift varying between -10 and -85 degrees.

The FRFs estimated with the various odd harmonic signals, listed as Tests 1 to 6 in Table 8-6, are shown as solid lines and the FRF estimated with the consecutive signal is shown as a dashed line. It is clear that they correspond very well, though there appears to be a very slight negative bias in the FRF estimated with the consecutive signal.

The accuracy of the estimated FRFs was investigated by calculating the standard deviation on the estimates, according to equation (2-35), for the 30 odd harmonic signal at the three input amplitudes. The amplitudes of the FRFs are shown in Figure 8-14, along with the 2σ bounds on the values. It can be seen that the uncertainty at $\pm$ 10% and $\pm$ 5% W_f is very small, whereas it does begin to become significant at $\pm$ 2% W_f. Plotting the three FRF amplitudes together, in Figure 8-14 (d), shows that they correspond very well and that the FRF does not depend on the amplitude of the input signal. The same pattern was found with the LP shaft, as shown in Figure 8-15.

The data gathered at an amplitude of $\pm$ 10% W_f will thus be used for parametric modelling, since they have the lowest uncertainty and the influence of any odd-order engine nonlinearities is clearly minimal for this input range. It is therefore interesting to compare the FRF estimated with these data to that estimated with the 15 quasi-log harmonics, which are to be used for the cross-validation of the parametric model.

The FRFs and their 2σ uncertainty bounds are plotted for the HP shaft in Figure 8-16 and the LP shaft in Figure 8-17. The FRFs show good agreement, though there is some discrepancy at higher frequencies, particularly in the phase, which is larger than the 2σ uncertainty bounds. The uncertainty of the FRF estimated with the 15 quasi-log harmonics is seen to be smaller than that with the 30 odd harmonics.

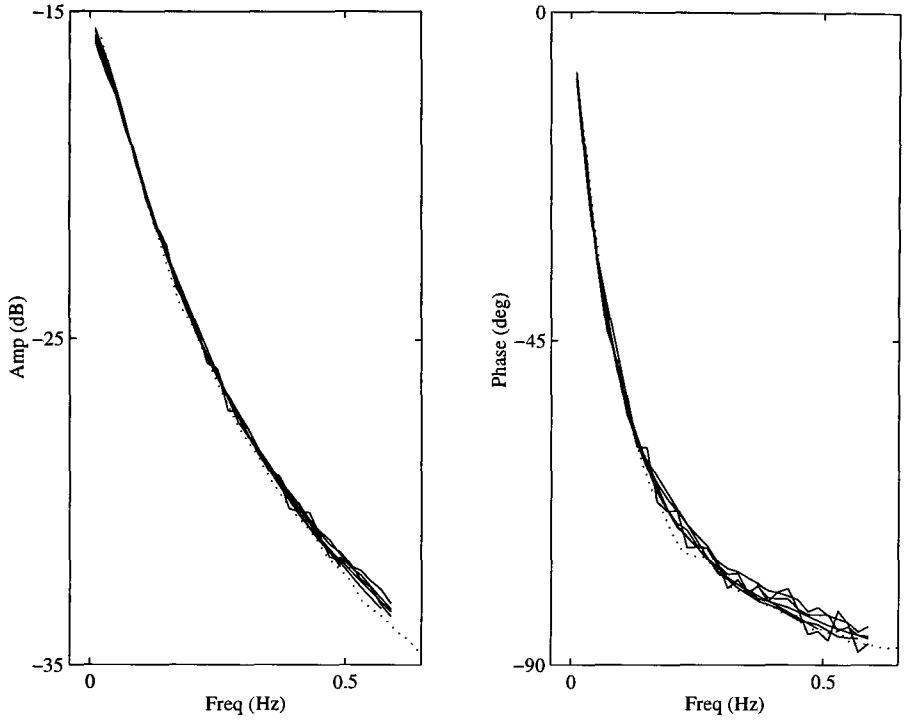

Figure 8-12. Estimated FRFs for HP shaft, Tests 1-6 (solid), Test 7 (dashed).

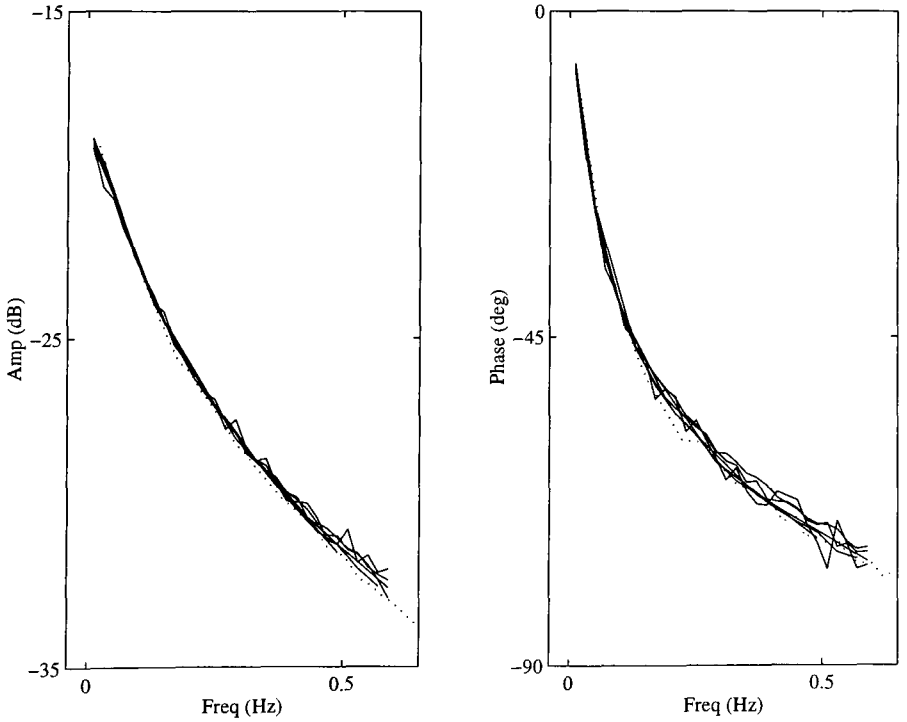

Figure 8-13. Estimated FRFs for LP shaft, Tests 1-6 (solid), Test 7 (dashed).

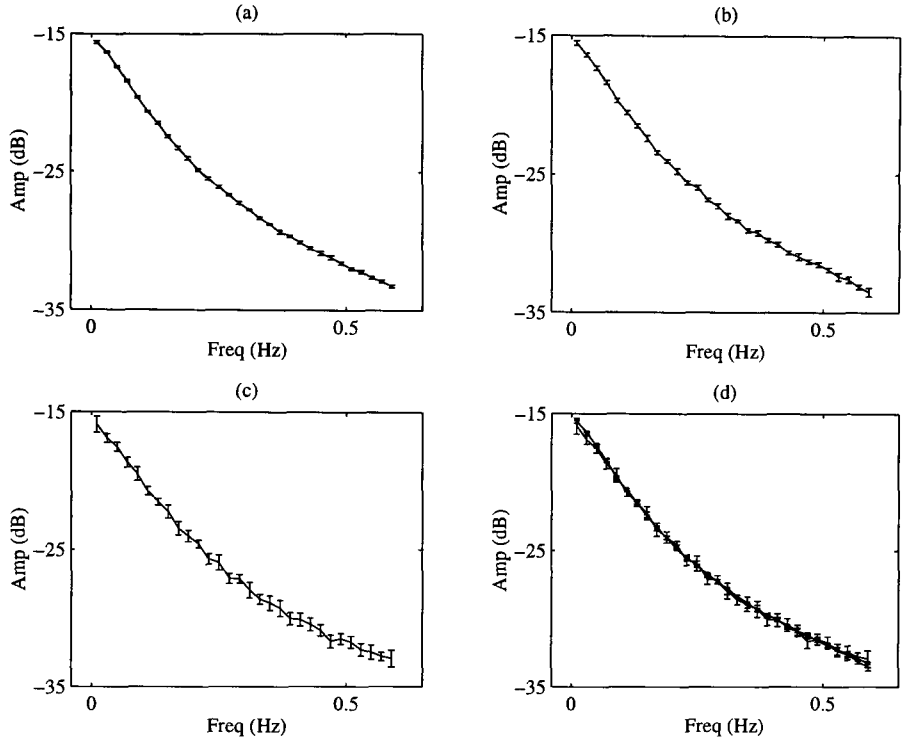

Figure 8-14. Amplitude of FRF estimates for HP shaft, using 30 odd harmonic data, showing 2σ bounds. With input amplitude (a) $\pm$ 10% W_f (b) $\pm$ 5% W_f (c) $\pm$ 2% W_f and (d) all plotted together.

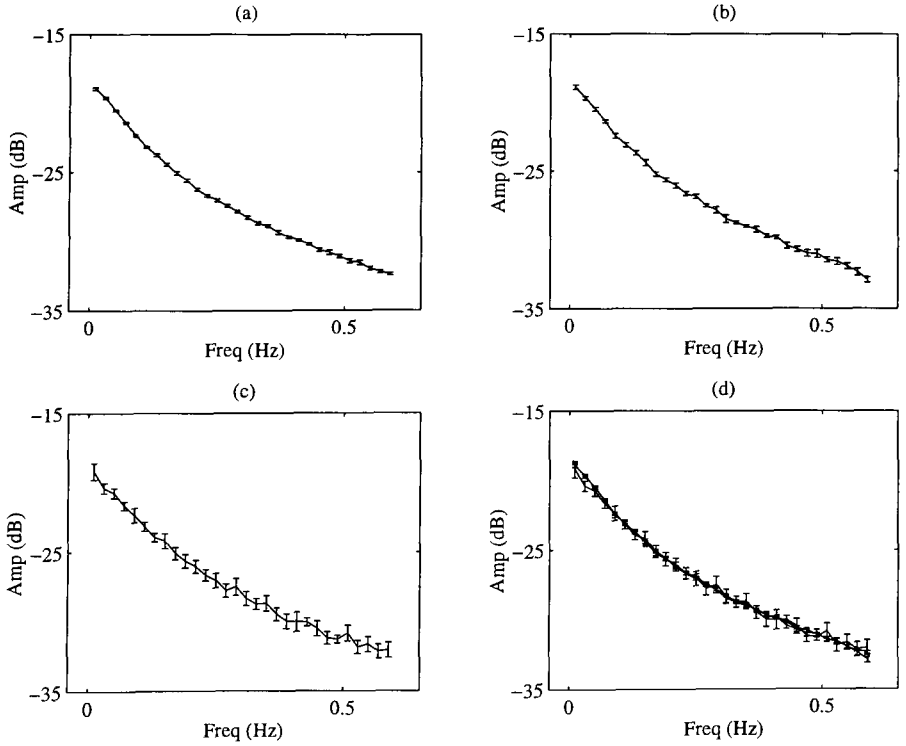

Figure 8-15. Amplitude of FRF estimates for LP shaft, using 30 odd harmonic data, showing 2σ bounds. With input amplitude (a) $\pm$ 10% W_f (b) $\pm$ 5% W_f (c) $\pm$ 2% W_f and (d) all plotted together.

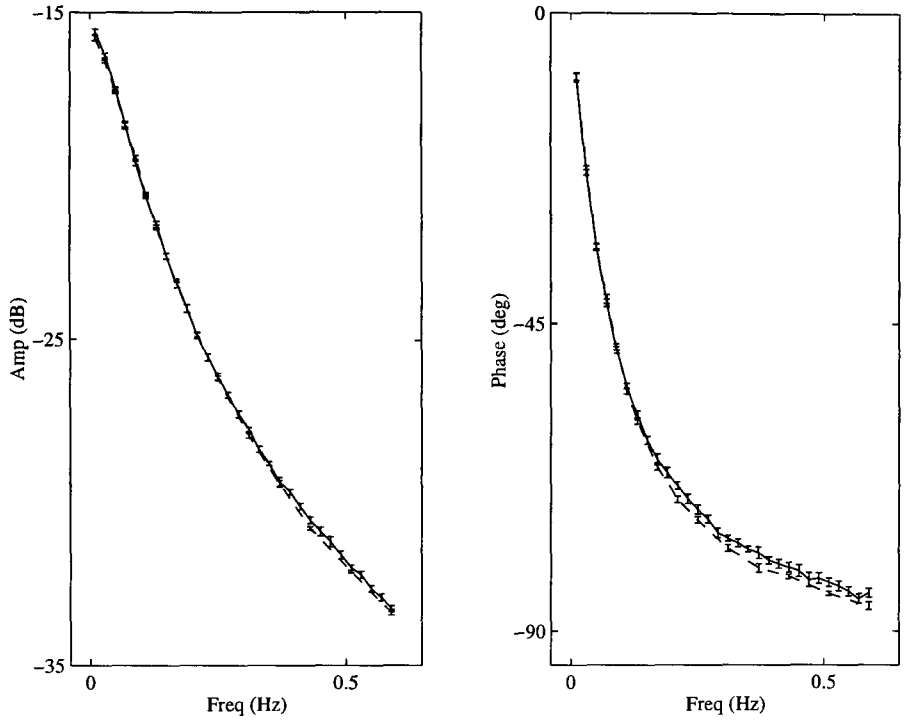

Figure 8-16. Estimated FRFs for HP shaft with 30 odd harmonic signal (solid) and 15 quasi log signal (dashed), showing 2σ bounds. Input amplitude $\pm$ 10% W_f.

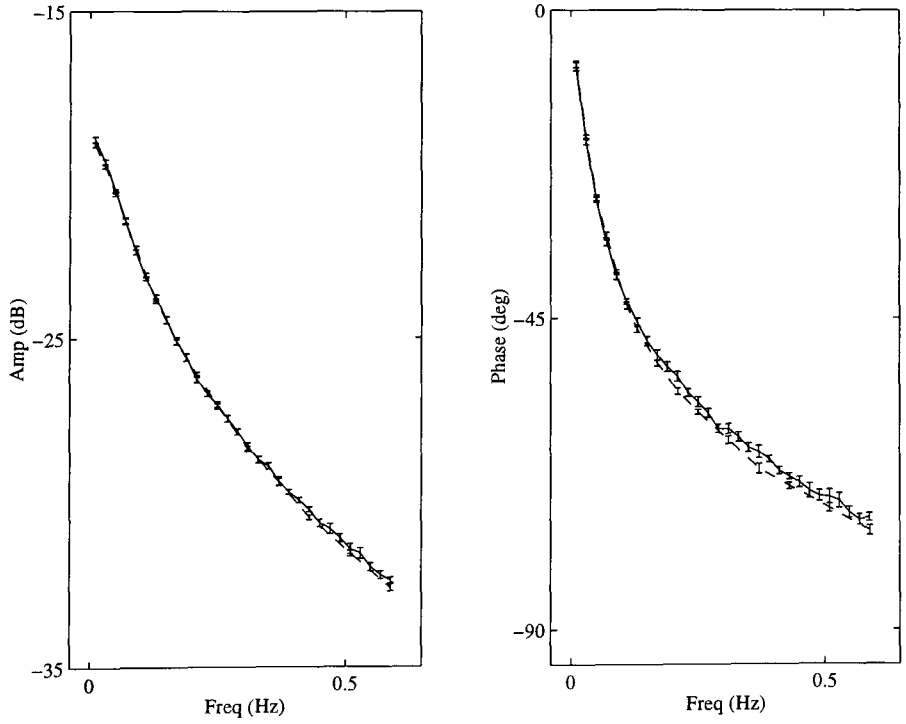

Figure 8-17. Estimated FRFs for LP shaft with 30 odd harmonic signal (solid) and 15 quasi log signal (dashed), showing 2σ bounds. Input amplitude $\pm$ 10% W_f.

8.5 Conclusions

The experimental set-up used to perform the engine tests has been described and the associated problems have been discussed. Data gathered during previous engine tests were used to assist in the design of a range of broad-band multisine test signals for the purpose of linear modelling, model cross-validation and nonlinear detection and modelling. It was concluded that the signals applied in previous tests were designed with too high a frequency content and the new designs were adjusted accordingly.

The data from previous tests also allowed the dynamics of the fuel feed system to be studied and no significant nonlinear behaviour was detected for an input amplitude range of $\pm$ 10% W_f. This allowed discrete linear models to be estimated, which were then transformed to the s-domain. The break-point of the lowest frequency fuel feed pole was found to be only ten times higher than that of the highest frequency pole of the engine. This confirmed the need to eliminate the fuel feed from the estimated engine model by using the measured fuel flow as the input.

Based on the information gained form the previous tests the bandwidth of the new range of signals was restricted to 0.01-0.6 Hz, which more than adequately covered the break-points of the engine poles and was also well below the break-points of the fuel-feed poles. The signals were successfully applied to the engine and a detailed analysis of the results was presented.

The time-domain analysis showed that the synchronisation between signal reconstruction and data sampling clocks was very good, that the influence of drift on the engine operating point was negligible and that the engine shaft speed responses showed good repeatability, when plotted period by period. The frequency-domain noise analysis revealed significant correlation between the input and output noise, particularly at low frequency, and the presence of peaks in the noise spectra. The frequency-domain SNRs after averaging were 40 dB or better at the input and the outputs.

The presence of nonlinear effects was investigated by calculating the nonlinear coherence of the odd-odd multisine at the unexcited frequencies. This showed the fuel feed system to be linear, with a weak even-order nonlinearity possibly present in the engine. The nonlinearity was not sufficiently significant to warrant an attempt at nonlinear modelling using the Volterra measurement multisine. The use of odd harmonic signals should thus eliminate any nonlinear effects at the test frequencies.

The FRFs were calculated for each test and found to be in good agreement for each of the odd harmonic signals. This was not the case for the consecutive harmonic signal, the FRF of which exhibited a slight bias. The accuracy of the estimated FRFs was assessed by calculating the statistical bounds on the estimates, which showed that the uncertainty was very small for input amplitudes of $\pm$ 5% W_f or greater. Thus an acceptable accuracy was achieved after averaging only six periods of the multisine signal. This resulted in a very significant reduction in engine test times, compared with testing using single sines (as described in Chapter 3).

In overall conclusion, high quality data have been gathered from the engine, which show good clock synchronisation and low drift. The influence of the weak even-order nonlinearity detected in the engine can be eliminated by the use of odd harmonic signals. Good SNRs were obtained, of 40 dB or better, and the accuracy and repeatability of the FRFs estimated with the different odd harmonic signals indicates that high quality parametric models can be estimated from these data. This shows that broad-band multisine signals can be applied with confidence to engine testing, in place of the more commonly used single sine tests, providing a considerable reduction in test times and consequently reducing costs.

Chapter IX

Gas Turbine Modelling

Abstract — *The results of the frequency-domain estimation of s-domain models of the HP and LP shaft dynamics are presented. The influence of a low-frequency effect and of the engine nonlinearity on the linear models is examined. The models are compared with the linearised thermodynamic model of the Spey engine at the same operating point. Discrete models are estimated in the time domain and compared to the frequency-domain models. The discrete models contain negative real poles and the possible causes of this effect are discussed and studied using simulations of the engine test set-up. Overall conclusions are drawn for the gas turbine modelling.*

9.1 Introduction

The frequency-domain estimation techniques discussed in Chapter 2 will now be applied to the gas turbine data analysed in Chapter 8. Frequency-domain estimation is well suited to this problem, since it allows the direct estimation of s-domain models, which are required to verify the thermodynamic engine models. The use of a measured input signal is assumed in the method employed and the nonparametric noise model of the estimator is particularly suited to the type of noise encountered in this system. The ability to include the pure time delay as an estimation parameter and the straightforward treatment of nonlinear effects are also attractive features of the frequency-domain approach.

Single-input, single-output models of the HP and LP shaft dynamics will be estimated using the 30 odd harmonic signal with an amplitude of ± 10% N_H and cross-validated using the 15 quasi-log harmonic signal of the same amplitude. The influence of engine nonlinearities will be examined by comparing the models estimated using the 30 odd and 30 even harmonic signals and also by analysing the frequency-domain residuals. The estimated models will be compared with the thermodynamic models and the models estimated by Hill (1994) at the same operating point.

Discrete models will also be estimated in the time domain and their poles and zeros compared with those of the frequency-domain models by means of the impulse invariant transform. It will be seen that the estimated models contain single real negative poles, which cannot be transformed to the s-domain and compared to the poles of the thermodynamic models. The possible causes of this effect will be discussed and studied using simulations of the engine test set-up. The influence of the sampling interval on time-domain estimation will also be examined, by progressively resampling the engine data at faster rates.

9.2 Frequency-Domain Estimation

The frequency-domain estimator proposed by Pintelon *et al.* (1992) will be employed to estimate transfer function models of the HP and LP shaft dynamics. The noise variances and covariance, which were calculated during the averaging of the frequency data, will be used as *a priori* information and the pure time delay will be included as a free parameter in each estimation. The expected value of the estimator cost function varies as a function of the number of estimation frequencies and the number of free parameters, as defined in equation (2-39).

9.2.1 High Pressure Shaft

The variation of the cost function with model order for the 30 odd harmonic signal is shown in Table 9-1. A pole and zero are added for each increase in model order, which represents the addition of a freely specified mode and modal gain. The major drop in the cost function occurs with the one zero, two pole (1/2) model, suggesting that this is the best structure. Table 9-2 shows the pole and zero positions of the first three of these models, with their standard deviations expressed as a percentage of their magnitudes.

The 1/2 model has a pole-zero pair close to the origin, whose 2σ regions of uncertainty do, however, overlap. Adding another pole and zero gives an unstable pole, which is not a credible model. The frequency responses of the 0/1 and 1/2 models are shown in Figures 9-1 and 9-2, where the additional pole and zero are seen to be modelling a localised low-frequency effect.

The large drop in the cost function between the 0/1 and 1/2 models appears to contradict the very small improvement in the model frequency fit shown in Figure 9-2. This can be explained by considering the estimator cost function, given in equation (2-37), where the noise variances appear in the denominator. This means that small values of noise variance will tend to increase the value of the cost function, making it more sensitive to small modelling errors.

In the case of the gas turbine data, the noise variances are indeed very small, as indicated by the SNRs listed in Table 8-9. This means that secondary effects must also be included in the model in order to bring the cost function close to its expected value, which is an indication of the high quality of the measured data.

TABLE 9-1

COST FUNCTION FOR HP SHAFT MODELS, 30 ODD HARMONIC SIGNAL

Order	Cost Function (K)	K_{min}
0/1	280	28.5
1/2	34	27.5
2/3	32	26.5
3/4	28	25.5

TABLE 9-2

ESTIMATION RESULTS FOR HP SHAFT, 30 ODD HARMONIC SIGNAL

Order	Delay (ms)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	8.5	—	—	-0.5015	0.2
1/2	13.2	-0.3405	9.2	-0.2749	7.3
				-0.6017	2.4
2/3	12.5	-0.3606	11	-0.2860	8.1
		-19.640	64	-0.6145	3.6
				16.336	39

These models were cross-validated using the 15 quasi-log harmonic data, the results for which are shown in Tables 9-3 and 9-4. They show the same pattern as the 30 odd harmonic results, with the major drop in the cost function occurring with the 1/2 model. This model has a dominant pole at s = -0.5 and a close pole-zero pair near the origin, the 2σ uncertainty regions of which once again overlap.

It could be argued that the close pole-zero pairs in the 1/2 models estimated with each data set are simply an indication of over-modelling and should be cancelled out. However, they may be highly correlated, such that they move together along the real axis, rather than cancel. They are also seen to model a real feature of the data in each case and the drop in the estimator cost function is very significant. They are clearly modelling a localised low frequency dynamic, though the difference between the models suggests that there is some uncertainty about the exact location of the pole and zero. The low frequency pole is at 0.044 Hz in the model estimated with the 30 odd harmonics and at 0.017 Hz in the model estimated with the 15 quasi-log harmonics. Since the fundamental frequency of the signals is only 0.01 Hz, the pole-zero pairs are clearly on the limits of modelling with these data.

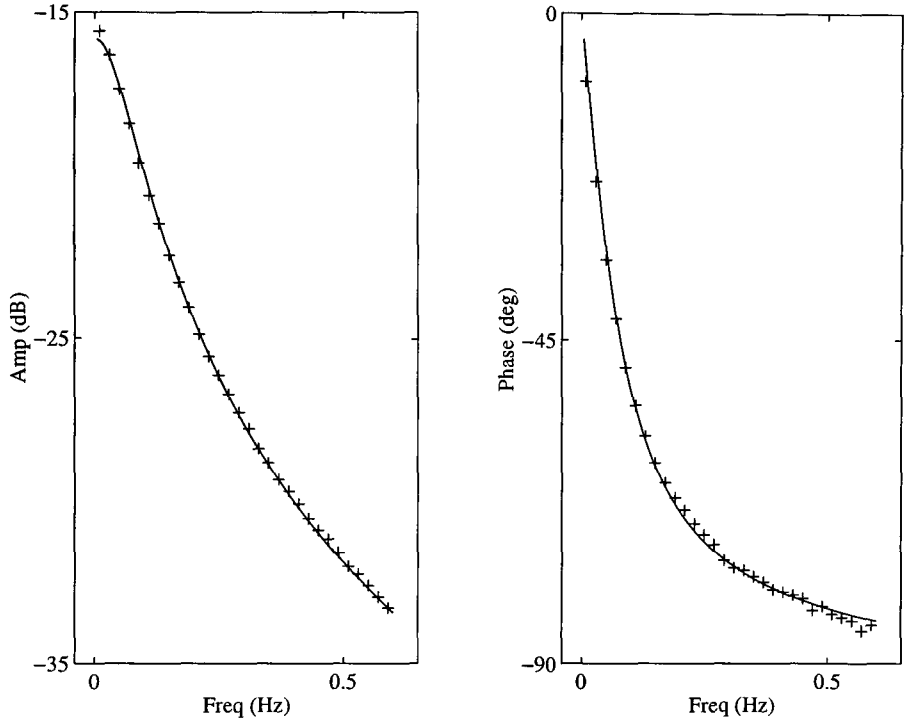

Figure 9-1. Frequency response of model (solid) and FRF (crosses), for 0/1 model, HP shaft.

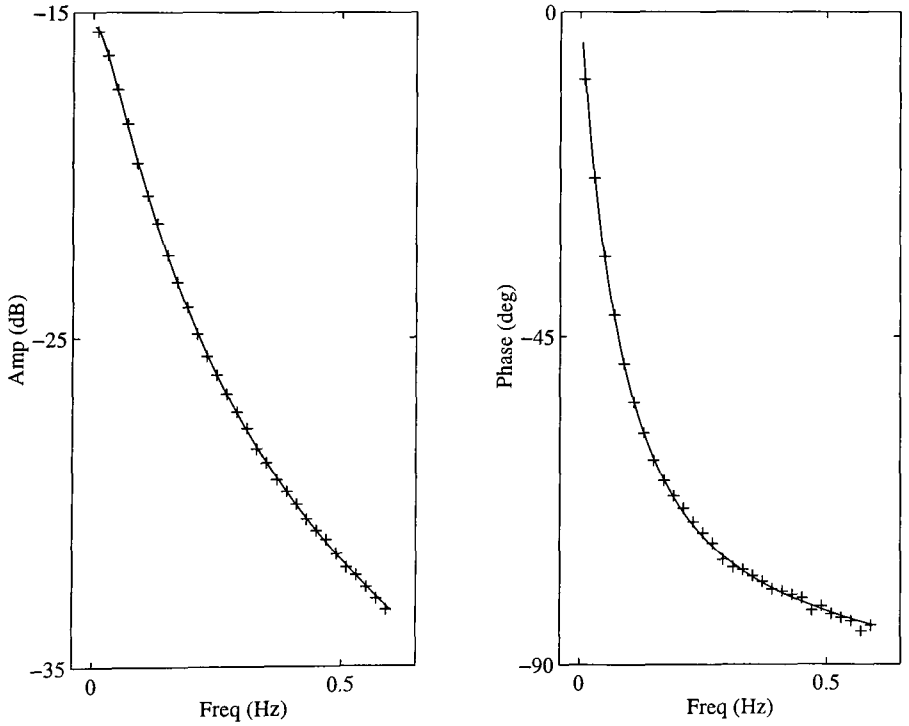

Figure 9-2. Frequency response of model (solid) and FRF (crosses), for 1/2 model, HP shaft.

TABLE 9-3

COST FUNCTION FOR HP SHAFT MODELS, 15 QUASI-LOG HARMONIC SIGNAL

Order	Cost Function (K)	K_{min}
0/1	146	13.5
1/2	36	12.5
2/3	34	11.5
3/4	30	10.5

TABLE 9-4

ESTIMATION RESULTS FOR HP SHAFT, 15 QUASI-LOG HARMONIC SIGNAL

Order	Delay (ms)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	17.3	—	—	-0.4963	0.2
1/2	18.6	-0.1140	16	-0.1054	16
				-0.5156	0.5
2/3	-105	-0.1317	16	-0.1212	16
		17.539	44	-0.5220	0.9
				-14.387	33

9.2.2 Low Pressure Shaft

The variation of the cost function for the LP shaft models, estimated using the 30 odd harmonic data, is presented in Table 9-5. The large drop in the cost function between the 0/1 and 1/2 models shows that the dynamics are at least second-order. There is even a case for selecting a 2/3 model, though the drop in the cost function is not as significant as that for the 1/2 model.

Table 9-6 shows that the 2/3 model has a very close pole-zero pair near the origin, the uncertainty regions of which do overlap. This close pole-zero pair is once again

modelling a low frequency effect in a similar way to the pole-zero pair in the 1/2 model of the HP shaft. Figure 9-3 shows the amplitudes of the model frequency responses and the difference between the 1/2 and 2/3 models is so small as to be almost indistinguishable.

TABLE 9-5

COST FUNCTION FOR LP SHAFT MODELS, 30 ODD HARMONIC SIGNAL

Order	Cost Function (K)	K_{min}
0/1	4609	28.5
1/2	69	27.5
2/3	39.55	26.5
3/4	39.54	25.5

TABLE 9-6

ESTIMATION RESULTS FOR LP SHAFT, 30 ODD HARMONIC SIGNAL

Order	Delay (ms)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	-32	—	—	-0.7939	0.2
1/2	14.7	-0.9064	2.2	-0.4199	1.2
				-1.8250	1.7
2/3	17.2	-0.1127	31	-0.1042	30
		-1.0496	3.9	-0.4754	3.3
				-1.9783	2.6
3/4	17.3	-0.1156	31	-0.1068	30
		-1.0533	4.1	-0.4768	3.4
		1.5033	28	1.5026	28
				-1.9831	2.7

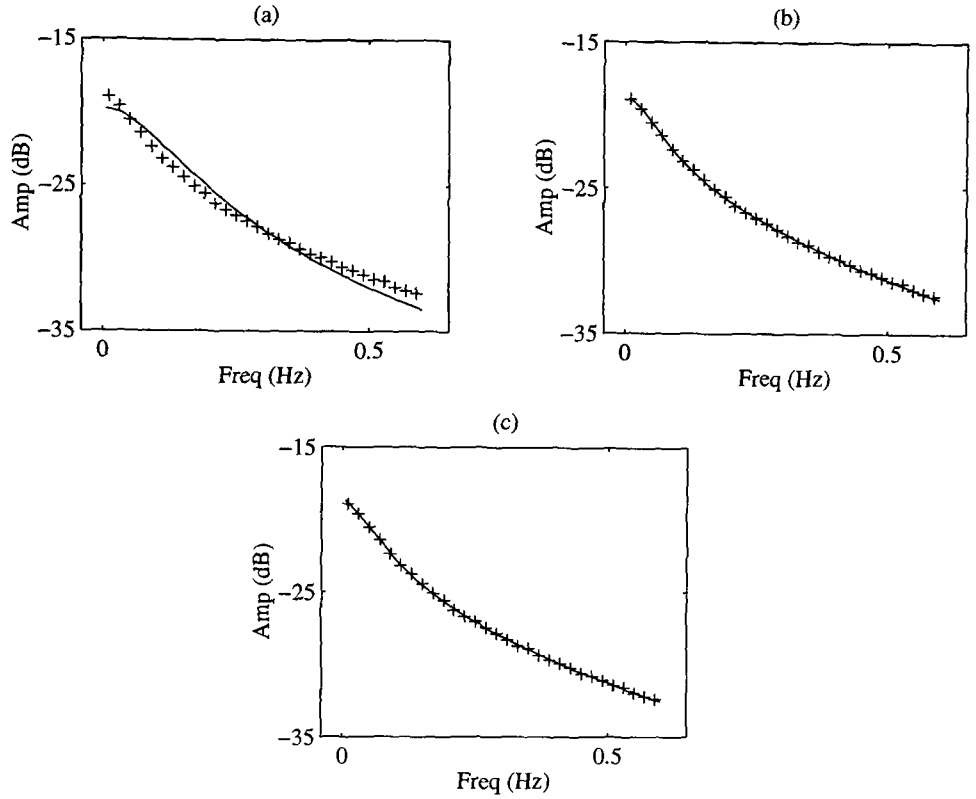

Figure 9-3. Amplitude response of LP shaft model (solid) and FRF (crosses), for model orders (a) 0/1, (b) 1/2 and (c) 2/3.

Once again, cross-validation was performed using the 15 quasi-log harmonics, the results for which are shown in Tables 9-7 and 9-8. There is a greater similarity between the two sets of models in this case than with the HP shaft. The low frequency pole-zero pairs are once again on the limits of modelling but there is a case for selecting the 2/3 model.

TABLE 9-7

COST FUNCTION FOR LP SHAFT MODELS, 15 QUASI-LOG HARMONIC SIGNAL

Order	Cost Function (K)	$K_{\min}$
0/1	3076	13.5
1/2	81	12.5
2/3	27	11.5
3/4	24	10.5

TABLE 9-8

ESTIMATION RESULTS FOR LP SHAFT, 15 QUASI-LOG HARMONIC SIGNAL

Order	Delay (ms)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	-35.6	—	—	-0.7380	0.2
1/2	15.9	-0.8808	2.7	-0.4352	1.3
				-1.6692	2.1
2/3	21.2	-0.1539	20	-0.1396	19
		-1.1981	5.5	-0.5412	3.8
				-2.0177	4.0
3/4	18	-0.1480	21	-0.1349	20
		0.9761	37	-0.5324	3.8
		-1.1583	5.5	0.9827	37
				-1.9555	4.0

9.2.3 Low Frequency Mode

It is clear that the HP and LP shafts have different order dynamics, with the HP shaft being predominantly first-order and the LP shaft second-order. In each case, the addition of a further pole-zero pair models a low frequency effect and significantly reduces the model cost function.

The influence of this low frequency mode on the estimated models can be assessed by excluding a number of the lower harmonics from the data set, re-estimating a range of models and monitoring the drop in the cost function as the model order is increased. This approach was applied to the 30 odd harmonic data for the HP shaft, with the lowest four, eight and then 12 harmonics progressively excluded from the data. The resulting drop in the cost function between the 0/1 and 1/2 models is shown in Table 9-9 and the influence on the estimated pole and zero positions shown in Table 9-10.

The role of the pole-zero pair in modelling a low frequency mode is clearly illustrated, since there is little benefit in increasing the model order to 1/2 once eight or more harmonics have been omitted from the data. Omitting harmonics has little influence on

the estimate of the dominant pole but has a significant influence on the pole-zero pair. A similar pattern was observed for the transition between the 1/2 and 2/3 models of the LP shaft.

Hence if a model is required that adequately describes the HP shaft speed dynamics in the frequency range 0.17 Hz - 0.59 Hz then the 0/1 structure will be sufficient. If a model is required which covers the complete frequency range of the test signals then the additional pole-zero pair should be included. Models can be estimated for selected frequency ranges in both the time and frequency domains but the ease with which this can be achieved in the frequency domain is an attractive feature of this approach.

TABLE 9-9

VARIATION OF COST FUNCTION WITH OMITTED FREQUENCIES

Omitted Frequencies	f_{min} (Hz)	Cost Function (K)	
		0/1 Model	1/2 Model
0	0.01	280	34
4	0.09	66	26
8	0.17	21.2	19.9
12	0.25	15.4	15.0

TABLE 9-10

VARIATION OF ESTIMATED MODELS WITH OMITTED FREQUENCIES

Omitted Frequencies	Zero	Pole 1	Pole 2
0	-0.3405	-0.2749	-0.6017
4	-0.3664	-0.2904	-0.6163
8	-0.5076	-0.4090	-0.6456
12	-51440	-34.29	-0.5423

9.2.4 Pure Time Delay

The time delay was included as a freely varying parameter in the frequency-domain estimation. Values of between 13 and 21 ms were obtained for the selected models on each shaft, which match the typical value of combustion delay proposed by Thompson (1974), which was discussed in Chapter 7. This small delay can present a problem in the time-domain estimation of discrete models, since faster sampling is required in order to exactly specify the delay. This can cause numerical problems, as the poles and zeros of the discrete models will tend to cluster around the point $z = 1$.

9.2.5 Influence of Engine Nonlinearity

The presence of a weak even-power nonlinearity was detected during the data analysis presented in Chapter 8. This even nonlinearity would be expected to influence the consecutive harmonic signal but not the odd-harmonic signals. A comparison will therefore be made between the HP shaft models estimated using the 30 odd harmonic and the 30 consecutive harmonic multisines, which have equal power and an approximately equal bandwidth. The model cost functions are presented in Table 9-11, along with their expected values and the 5% and 95% χ^2 bounds on those values.

TABLE 9-11

COST FUNCTION FOR HP SHAFT MODELS

Order	K		K_{min}		
	Consecutive	Odd	5% <	χ^2	< 95%
0/1	257	280	20 <	28.5	< 38
1/2	228	34	19 <	27.5	< 37
2/3	215	32	18 <	26.5	< 35

The cost function of the 1/2 model estimated using the odd harmonic data falls below the upper bound on the theoretical minimum, while it remains much larger for the consecutive data. Increasing the model order still further is of no real benefit in the consecutive case, which strongly suggests the presence of unmodelled nonlinear dynamics in the data. The estimated zeros, poles and pure time delay for the 0/1 and 1/2 models obtained with each signal are shown in Table 9-12.

TABLE 9-12

ESTIMATION RESULTS FOR HP SHAFT MODELS

Signal	Order	Delay (ms)	Zeros	σ_Z (%)	Poles	σ_P (%)
Consec	0/1	21.1	—	—	-0.4826	0.3
	1/2	14.5			-0.4908	0.4
			-4.1361	15	-3.9343	14
Odd	0/1	8.5	—	—	-0.5015	0.2
	1/2	13.2	-0.3405	9.2	-0.2749	7.3
					-0.6017	2.4

For both first- and second-order structures, the difference between the models estimated with each signal is greater than the 2σ confidence intervals. This is particularly true of the 1/2 model estimated with the consecutive harmonic signal, where the pole-zero pair is placed at a much higher frequency. It appears that the even nonlinearity is masking the low frequency effects in the consecutive signal, which are however detected by the odd harmonic signals.

Analysis of Residuals. The presence of unmodelled nonlinear dynamics can be detected using the normalised autocorrelation of the frequency-domain residuals, defined in equation (5-15) and denoted R_{ee}. The R_{ee} of the 0/1 and 1/2 HP shaft models estimated using the consecutive signal are plotted in Figure 9-4 and for the odd harmonic signal in Figure 9-5. The value of R_{ee} for the consecutive data is large at a lag of zero and is unaffected by increasing the model order. This suggests that the main contribution to the modelling errors comes from unmodelled nonlinear dynamics. For the odd data, R_{ee} is initially broad in shape, with large values at small lags. Increasing the model order has a significant effect on its value, leaving only a small peak at a lag of zero. This indicates the presence of unmodelled linear dynamics in the 0/1 model, which are eliminated by increasing the model order. This is further confirmation that nonlinear effects are present in the consecutive data and also that a 1/2 structure is required to adequately model the HP dynamics.

The variation of R_{ee} with model order for the LP shaft is shown in Figures 9-6 and 9-7. It has a broad shape for both 0/1 models, which suggests the presence of unmodelled linear dynamics in the consecutive and odd harmonic data. Increasing the model order to 1/2 produces an R_{ee} with a sharp peak at a lag of zero in both cases, however the value of the peak is much smaller with the odd harmonic data. A further increase in the model order reduces the magnitude of the peak still further with the odd data but has no effect with the consecutive data.

This once again illustrates the influence of the engine nonlinearity on the consecutive harmonic data and suggests that the LP dynamics are at least second-order, if not third-order. In this case, the choice of model order is not as clear as it was with the HP shaft, since increasing the order from 1/2 to 2/3 does not have such a significant effect on R_{ee}.

Since the aim of the gas turbine modelling is to produce high quality linear models, in order to verify the established thermodynamic models, it is important to assess the influence of any nonlinear effects. The techniques outlined in this thesis allowed the detection of a weak even power nonlinearity, which in this case has a significant influence on the 1/2 model estimated with consecutive signal data. This effect is eliminated using odd harmonic signals.

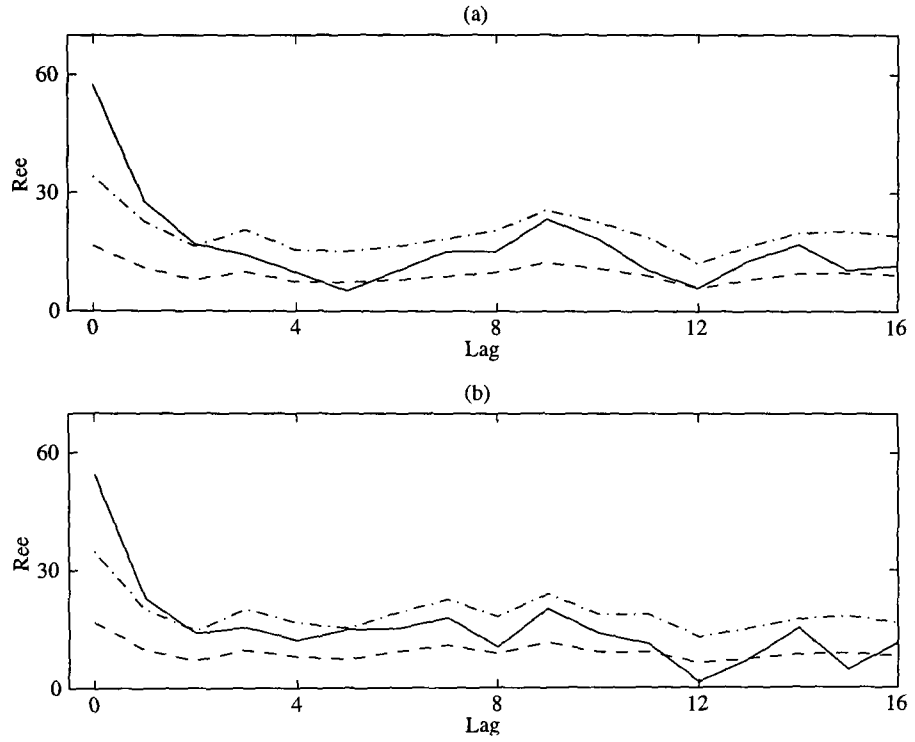

Figure 9-4. R_{ee} for consecutive multisine, HP shaft data: (a) 0/1 model and (b) 1/2 model, with 50% (dashed) and 95% (dashdot) confidence intervals.

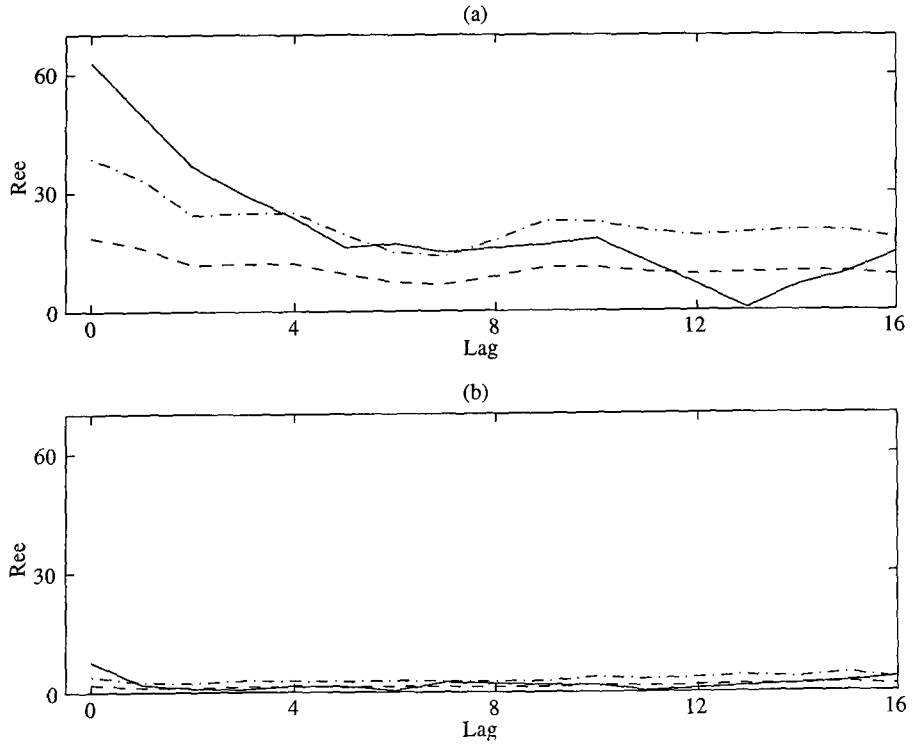

Figure 9-5. R_{ee} for odd multisine, HP shaft data: (a) 0/1 model and (b) 1/2 model, with 50% (dashed) and 95% (dashdot) confidence intervals.

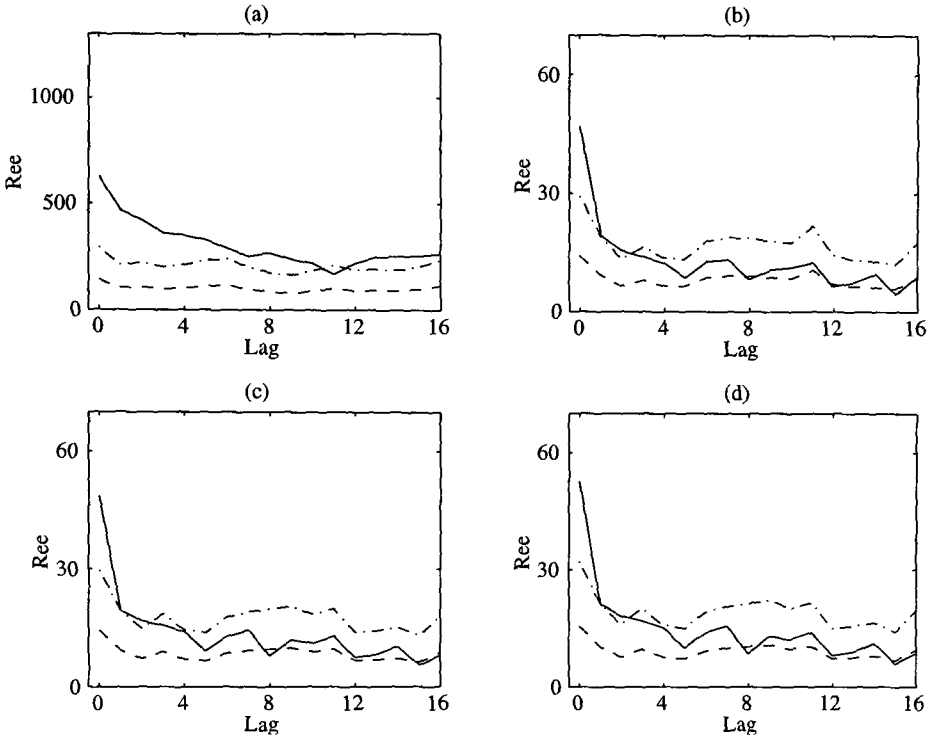

Figure 9-6. R_{ee} for consecutive multisine, LP shaft data: (a) 0/1, (b) 1/2, (c) 2/3 and (d) 3/4 models, with 50% (dashed) and 95% (dashdot) confidence intervals.

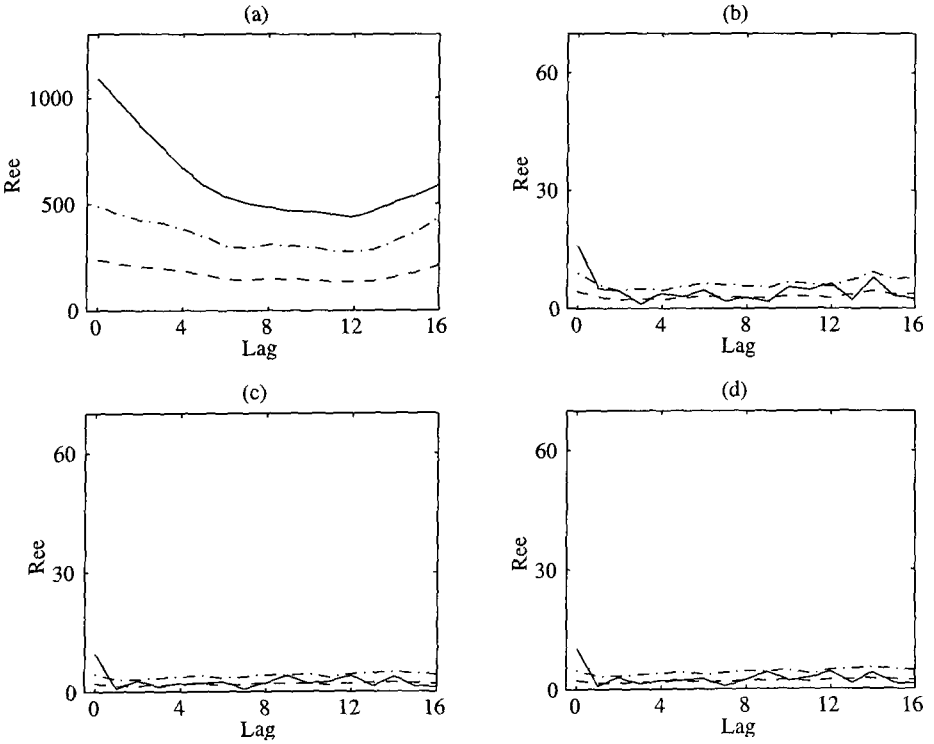

Figure 9-7. R_{ee} for odd multisine, LP shaft data: (a) 0/1, (b) 1/2, (c) 2/3 and (d) 3/4 models, with 50% (dashed) and 95% (dashdot) confidence intervals.

9.2.6 Comparison with Thermodynamic Models

Models of the HP and LP shaft have been estimated with a high degree of confidence from the test data and it is now possible to compare these with the thermodynamic models derived from the engine physics. The transfer function matrix of the thermodynamic models can be calculated using the relation

$$\mathbf{G}(s) = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B} + \mathbf{D} \qquad (9-1)$$

In Table 9-13, the poles and zero of the HP shaft model, estimated with the 30 odd harmonic data, are compared to those of the thermodynamic model. Each of the models has a dominant pole close to the same point (around 0.1 Hz) and in addition a close pole-zero pair, which indicates that the HP shaft dynamics are predominantly first-order. The major difference is that the pole-zero pair of the estimated model is much closer to the origin. This has been shown to model a real feature of the data but it is clearly modelling a different effect from that modelled by the higher frequency pole-zero pair in the thermodynamic model.

TABLE 9-13

COMPARISON OF MODELS, HP SHAFT

	Estimated Model	Thermodynamic Model
Zeros	-0.34	-3.14
Poles	-0.60	-0.70
	-0.27	-3.16

For the LP shaft, the 1/2 thermodynamic model is compared to the 2/3 model estimated using the 30 odd harmonic data in Table 9-14. The separation of the zero and poles of the thermodynamic model is greater than with the HP shaft, which suggests that the dynamics are at least second-order.

Neglecting for a moment the close pole-zero pair in the estimated model, a direct comparison can be made between the models. The dominant pole of the estimated model is at a lower frequency than that of the thermodynamic model, at a frequency of 0.08 Hz

compared to 0.1 Hz. The zero and second pole are also at a lower frequency, which suggests that the LP shaft response is actually different from that predicted by the thermodynamic model.

TABLE 9-14

COMPARISON OF MODELS, LP SHAFT

	Estimated Model	Thermodynamic Model
Zeros	-1.05	-2.75
	-0.11	
Poles	-0.48	-0.70
	-1.98	-3.16
	-0.10	

It is important to keep in mind that different types of models are being compared, since the thermodynamic models are converted from a state space form, while the estimated models were directly estimated as transfer functions. The state space structure imposes common eigen values on the HP and LP shafts and hence common poles in the converted models. This is not the case for the frequency-domain transfer function estimator, which has greater freedom to fit poles exactly to the data. This is an area which warrants further study in the future.

Comparing the estimated models of the HP and LP shafts it is seen that they share a close, but not identical, dominant mode at a frequency of around 0.09 Hz. The results suggest that the HP *shaft dynamics* are first-order and the LP *shaft dynamics* second-order. The additional pole-zero pairs are too slow to be associated with these shaft dynamics and are probably modelling a heat soakage effect.

As has been discussed in Chapter 7, such effects become prominent during large transient manoeuvres of an engine and while the tests were carried out around a fixed operating point there may have been sufficient variation of the engine operation during the tests to excite some heat transfer dynamics. It is possible that the low frequency modes are modelling one of the faster heat soakage effects, such as blade expansion. Blade expansion time constants of 5 seconds were quoted by Hennecke and Trappman (1982)

and blade and casing time constants of 2 seconds and 10 seconds respectively, by Pilidis and Maccallum (1982). The estimated low frequency modes have time constants of between 3 and 10 seconds.

9.2.7 Summary

The frequency-domain estimation results indicate that:

- High quality models of the shaft speeds can be estimated using frequency-domain techniques.
- The HP shaft dynamics are first-order and the LP shaft dynamics are second-order. The shafts share a close, but not identical, dominant mode.
- The model fits can be further improved by the addition of low frequency pole-zero pairs, which are thought to be modelling a fast heat soakage effect, such as blade expansion.
- The pure time delay is around 15 ms, which matches previous estimates for typical values of combustion delay.
- The quality of the estimated linear models is improved by eliminating the influence of the even-order engine nonlinearity, achieved by using odd harmonic multisines.
- Comparison with the thermodynamic models shows a reasonably good match for the dominant modes of each shaft but suggests that the second-order dynamics of LP shaft are different from those predicted by the thermodynamic model.

9.3 Time-Domain Estimation

As discussed in Chapter 7, recent work by Hill (1994, 1997) examined the time-domain identification of discrete engine models. Each of the second-order transfer function models estimated in this way contained a negative real pole, which could not be transformed to the s-domain to allow comparison with the thermodynamic models. It was therefore of interest to see if this effect could be replicated with these data.

Discrete *output error* (OE) models were estimated in the time domain, using the complete data sets of the 30 odd harmonic and 15 quasi-log harmonic signals. These consisted of 3,000 samples of the measured input and output signals, which were recorded at a

sampling frequency of 5 Hz, with their means removed. The use of periodic signals allowed the SNRs to be improved by calculating the FFT of each data record, setting the values of all nonexcited frequency lines to zero, calculating the inverse FFT and using these new data sets in the estimation.

9.3.1 High Pressure Shaft

The HP shaft models estimated with the 30 odd harmonic signal are presented in Table 9-15. The major drop in the cost function occurs with the 1/2 model, which matches the results obtained with frequency-domain estimation. Figure 9-8 shows the frequency response of the 0/1 model, the amplitude of which has a reasonably good fit to the estimated FRF. The phase response, on the other hand, is very different to that of the FRF and appears to be heading far beyond the maximum shift of -90° expected with a first-order model.

This effect can be largely attributed to the slow sampling frequency of the data, relative to the dynamics of the continuous system. This has little effect on the amplitude response of the discrete model but introduces a significant additional phase lag (see Franklin *et al.* (1990), pp. 193-197). The effect of resampling the data at higher frequencies will be discussed in Section 9.3.4. The frequency response of the 1/2 model is shown in Figure 9-9.

The time response of the 0/1 model is shown in Figure 9-10, where it is seen to match closely the measured output. This match is further improved with the 1/2 model, as shown in Figure 9-11. The input-output properties of the discrete models are thus very good.

TABLE 9-15

TIME-DOMAIN ESTIMATION RESULTS FOR HP SHAFT, 30 ODD HARMONIC SIGNAL

Order	Cost Fn. (V e-4)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	168	—	—	0.9003	0.02
1/2	13	0.0225	10	0.9094	0.01
				-0.9936	0.6
2/3	3	0.9625	0.05	0.9662	0.04
		0.0195	17	0.9005	0.02
				-0.9991	1.1

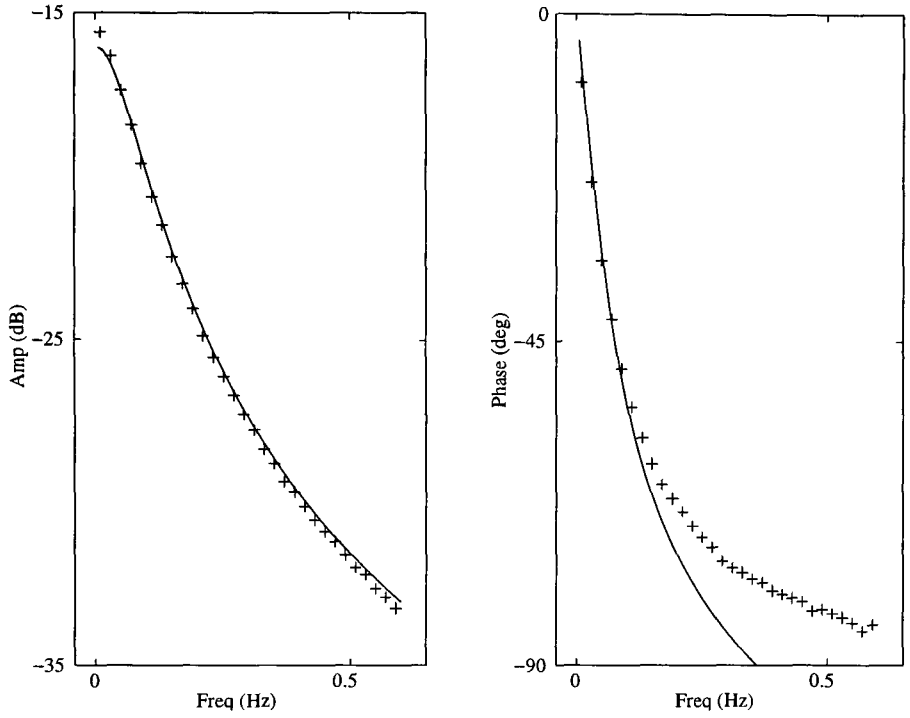

Figure 9-8. Frequency response of discrete model (solid) and FRF (crosses), for 0/1 model, HP shaft.

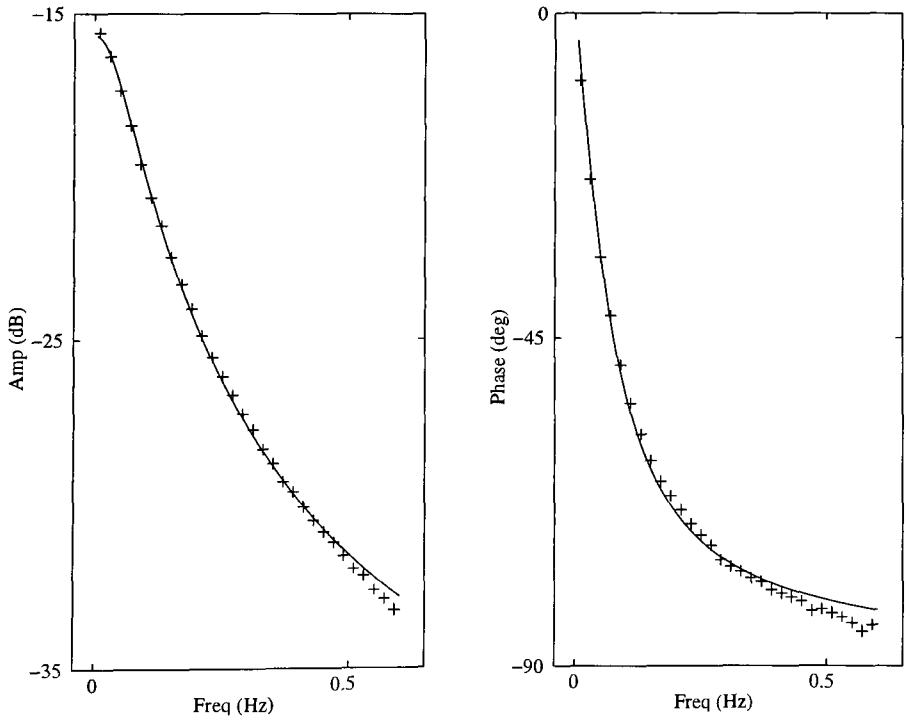

Figure 9-9. Frequency response of discrete model (solid) and FRF (crosses), for 1/2 model, HP shaft.

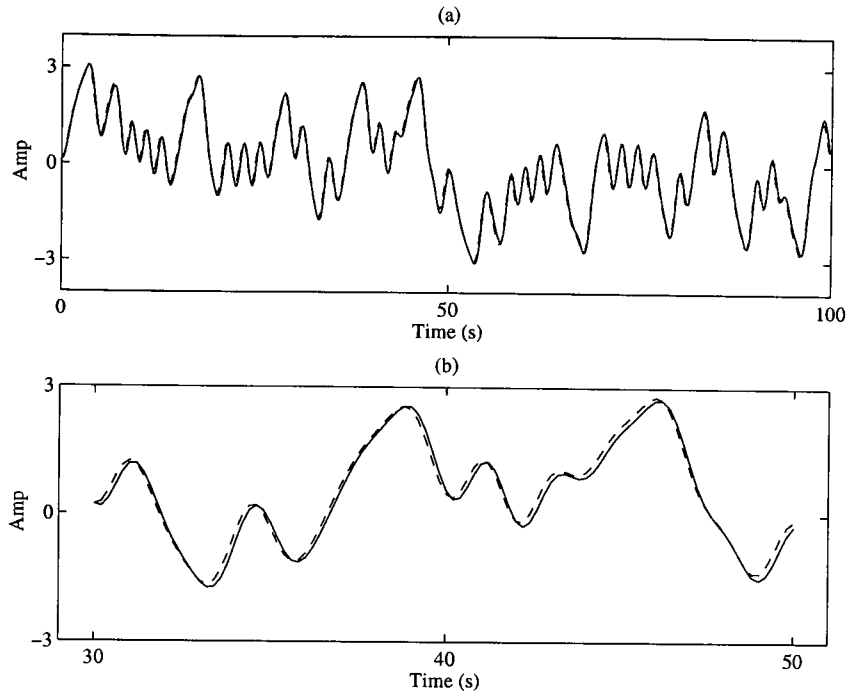

Figure 9-10. Discrete model output (solid) and measured engine output (dotted) for (a) one period and (b) a portion of that period, for 0/1 model, HP shaft.

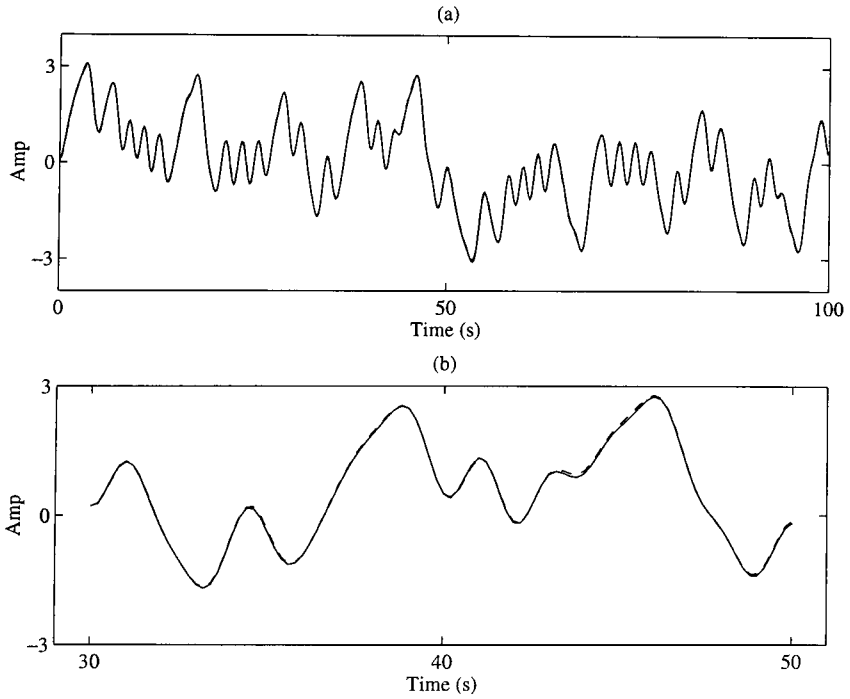

Figure 9-11. Discrete model output (solid) and measured engine output (dotted) for (a) one period and (b) a portion of that period, for 1/2 model, HP shaft.

The cross-validation results using the quasi-log data are listed in Table 9-16 and show a similar pattern to those of the 30 odd harmonic data, since the major drop in the cost function occurs with the 1/2 model. The significant feature of the 1/2 models, estimated with both the 30 odd harmonic signal and the 15 quasi-log harmonic signal, is the presence of a single, real, negative pole close to the unit circle. This will be referred to simply as the *negative pole* in the discussion which follows. Each of the models also has a zero close to the origin.

This means that the discrete models are not equivalent to the continuous models estimated in the frequency-domain. The impulse invariant transform of the second-order continuous models listed in Tables 9-2 and 9-4 yields discrete poles and zeros in the region of $z = 0.88$ to 0.98. This result is clearly not replicated in the models listed in Tables 9-15 and 9-16.

A range of other discrete model structures was also estimated on these data, which included ARX, ARMAX and BJ models, and in each case the 1/2 model was found to be the most appropriate (Ljung, 1994). These models also contained a negative pole and a zero close to the origin, which indicates that the problem is not simply due to the lack of a noise model in the OE structure.

TABLE 9-16

TIME-DOMAIN ESTIMATION RESULTS FOR HP SHAFT, 15 QUASI-LOG HARMONIC SIGNAL

Order	Cost Fn. (V e-4)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	78	—	—	0.9006	0.01
1/2	11	-0.0452	5	0.9077	0.01
				-0.9965	0.3
2/3	3	0.9819	0.02	0.9832	0.02
		-0.0962	1.3	0.9030	0.001
				-0.9993	0.05

9.3.2 Low Pressure Shaft

The estimation results for the LP shaft are shown in Table 9-17 for the 30 odd harmonic signal and in Table 9-18 for the 15 quasi-log harmonic signal. In the case of complex poles and zeros, Table 9-17 lists the worst case standard deviation of either the real or imaginary part. In a similar manner to the frequency-domain LP shaft results, the largest drop in the cost function occurs with the 1/2 models. However, there is a further significant reduction with the 2/3 models, which suggests that they should be considered. The 2/3 LP shaft models share common features with the HP shaft 1/2 models: the presence of a negative pole and a zero close to the origin.

TABLE 9-17

TIME-DOMAIN ESTIMATION RESULTS FOR LP SHAFT, 30 ODD HARMONIC SIGNAL

Order	Cost Fn. (V e-4)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	375	—	—	0.8387	0.07
1/2	24	0.6532	0.2	0.8996	0.02
				0.2285	1.4
2/3	2.3	0.8747	0.05	0.9288	0.02
		0.0693	1.9	0.7623	0.08
				-0.9994	0.07
3/4	2.1	0.8597	0.06	0.9255	0.01
		$-0.25 \pm j0.33$	10	0.7326	0.11
				$-0.97 \pm j0.20$	20

TABLE 9-18

TIME-DOMAIN ESTIMATION RESULTS FOR LP SHAFT, 15 QUASI-LOG HARMONIC SIGNAL

Order	Cost Fn. (V e-4)	Zeros	σ_Z (%)	Poles	σ_P (%)
0/1	213	—	—	0.8539	0.05
1/2	17	0.6789	0.3	0.8995	0.02
				0.3451	1.0
2/3	2.4	0.8892	0.06	0.9298	0.02
		0.0721	3.5	0.7905	0.08
				-0.9864	0.7
3/4	1.5	0.9445	0.05	0.9536	0.03
		0.7464	0.22	0.8757	0.06
		-0.0338	4.9	0.6558	0.29
				-1.0000	0.11

9.3.3 The Single Real Negative Pole

The 1/2 HP shaft models and the 2/3 LP shaft models each have a negative pole. If the impulse invariant transform is applied to such a pole then a single complex s-domain pole will result, since $s = \log_e(z)\, f_s$. Such models do not correspond to a physically realisable system, since a continuous system with a single complex pole will generate a complex output when subjected to a real valued input.

Kollár (1996) has argued that it is possible to transform such a pole, based on the observation that the mapping of s-domain complex conjugate poles to the z-domain converges to a single point on the negative real axis as the sampling frequency is reduced. This corresponds to a point at which the system poles are on the margin of aliasing. The method proposed by Kollár assigns a complex s-domain pole pair to the system, at the Nyquist frequency. Since the poles are aliased there is a whole series of s-domain models which are equivalent to the z-domain model at the sampling instants but which differ in their intersample behaviour. Kollár then selects the model which minimises the intersample oscillations.

This technique was applied to the 1/2 HP shaft model listed in Table 9-15, and the equivalent transform is shown in Table 9-19, along with the s-domain model directly estimated in the frequency domain on the same data. Since the additional poles and zeros are placed close to the Nyquist frequency they are completely different from those estimated in the frequency domain and also completely outside the frequency range of the test signals employed. For these reasons it is felt that application of this technique is not appropriate if a physical interpretation is to be made of the s-domain model, since the additional poles and zeros are clearly not modelling physical effects.

TABLE 9-19

ESTIMATED FREQUENCY-DOMAIN MODEL AND KOLLAR EQUIVALENT

Model	Zeros	in Hz	Poles	in Hz
Frequency	-0.34	0.05	-0.60	0.1
domain			-0.27	0.04
Kollár	$-13.51 \pm j8.72$	2.56	-0.47	0.08
Equivalent			$-0.03 \pm j15.71$	2.5

This means that, for the purposes of this work, discrete models with a pole which has no continuous counterpart have been estimated. It is not possible to compare this pole to those of the s-domain thermodynamic models, or make a physical interpretation of its significance. The generation of negative poles was found to occur with all the 1/2 models of the HP shaft and 2/3 models of the LP shaft estimated with data from Tests 1 to 5. A range of possible causes have been suggested for this effect, which will now be discussed.

Influence of Noise. The frequency-domain analysis of the test data showed the presence of significant correlation between the input and output noise, which can cause problems in time-domain estimation. The whiteness of the residuals of the OE models was checked by calculating their autocorrelation. Figure 9-12 shows the autocorrelation of the residuals for the 1/2 model of the HP shaft, estimated using the 30 odd harmonic data, which has significant values at non-zero lags.

Estimating an ARMAX or BJ structure considerably improved the whiteness of the residuals but did not eliminate the negative pole. The inclusion of a noise model did, however, influence its position, with the pole appeared at around $z = -0.5$ with the

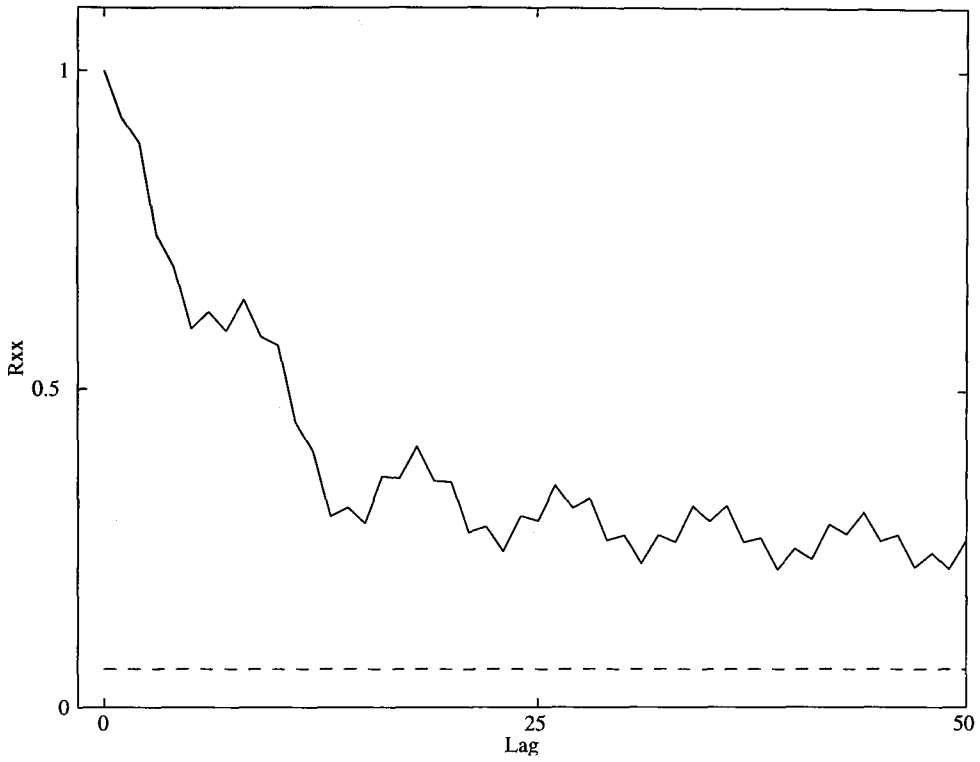

Figure 9-12. Normalised autocorrelation of the residuals (solid), with 99% confidence interval (dashed), for the 1/2 HP shaft model, 30 odd harmonic data.

ARMAX models and around z = -0.2 with the BJ structures. These results replicate the problems encountered by Hill (1994), who also estimated ARMAX models, using quite different input signals and a different sampling frequency.

Negative poles can also be generated when the SNR is very poor, causing pole-zero pairs to "jump" around within the unit circle. This effect occurred in estimations performed by Hill (1994) on data from noisy simulations of the Spey thermodynamic models. However, the SNR has to be very poor for this effect to occur and the SNR of the engine data was in all cases better than 40 dB.

Crude Quantisation. Åström (1980) noted that negative poles or second-order oscillatory modes have frequently been found when estimating models from data with crude quantisation. This is explained by the round-off introducing an oscillatory behaviour into the data.

Model Errors Due to Nonlinea***rity.*** It is possible that the modelling errors introduced by nonlinear effects might lead to the estimation of negative poles. This explanation can be discounted in this case, since the influence of the engine nonlinearity on the odd harmonic signals has been shown to be negligible.

Pure Time Delay. The time delay, which is less than one sampling period, may be causing the estimator to approximate the delay effect by generating models with negative poles and zeros close to the origin. Pure time delays of around 20 ms were estimated using the frequency-domain techniques, whereas the data sampling period was 200 ms. The use of a real negative zero to approximate the influence of a delay smaller than one sampling period is a standard technique (Franklin *et al.* 1990, pp. 44-46).

Aliasing of System Poles. This effect has been discussed previously, in relation to the Kollár (1996) transform technique, and while this is certainly an attractive explanation for the effect encountered it is unlikely to be the true cause of this problem, since the same effect was obtained by Hill with data sampled at a much faster rate.

Over-modelling. Under conditions of over-modelling, the locations of the poles and zeros which are not necessary to model the system dynamics are very inexact (Söderström (1975) and Hollkamp and Batill (1992)). It is possible that this is the source of the problem, since it is clear that the poles and zeros in question are on the limits of modelling with these data. However, under conditions of over-modelling the surplus poles and zeros tend to migrate together, as pairs, which is not the case with the present models.

Violation of ZOH Assumption. If a physical interpretation is to be made of a discrete model, the input signal must be applied to the system under ZOH conditions. In these tests, the fuel feed band-limits the input signal before sampling and hence this assumption is violated. It has been shown in simulation by Schoukens *et al.* (1994a) that violating the ZOH assumption can produce negative poles under certain conditions.

Model Sensitivity. This explanation was proposed by Hill, who showed that the z-domain poles of transfer function models are sensitive to errors in the parameters, especially for higher order models and models estimated at a high sampling rate. The shaft dynamics have two poles of similar magnitude, which makes them particularly

sensitive to parameter errors. Hill showed that state-space models have a much smaller sensitivity, if the model states are carefully chosen, and found that the estimated state space models did not contain negative poles.

Each of the above explanations attempts to address the cause of the modelling error, except the last, which addresses the effect of parameter errors on the poles and zeros of the model. In discussing this problem, it is important to distinguish between the cause of the error and its effect. Hill identified the closeness of the engine poles and the sensitivity of the z-domain models as the cause of the problem, but it is felt that this only addresses half the issue. The reason that the estimated model parameters are in error in the first place has not been adequately discussed. An explanation needs to be found for the source of the error, before turning to Hill's work for an understanding of how this error in the transfer function parameters can produce such a severe effect as a negative real pole.

Of the explanations discussed, only those relating to the correlated noise, the pure time delay, over-modelling and the violation of the ZOH assumption are felt to be plausible in this case. These will now be studied through simulation.

Simulation Study. By conducting a linear, noise-free simulation the problems of correlated noise and over-modelling can be immediately eliminated. This leaves the effect of the time delay and the violation of the ZOH assumption to be investigated. A block diagram of the overall simulation set-up is shown in Figure 9-13. The fuel feed system was the s-domain equivalent of the discrete model presented in Chapter 8, with the pure time delay removed, and the engine model was the 1/2 HP shaft model estimated using the 30 odd harmonic signal. The pure time delay was set to a nominal value of 20 ms, based on the estimates obtained in the frequency domain, which were discussed in Section 9.2.4. The sampling interval was set equal to that employed in the tests, of 200 ms, so that the time delay was less than one sampling interval.

The 30 odd harmonic signal was used as an input signal and a continuous simulation was performed, employing a simulation step size 3,000 times smaller than the fastest engine time constant, in order to ensure accuracy. The period of the ZOH was set equal to 200 ms and the data decimated after simulation, in order to arrive at the same sampling interval and number of samples as returned in the tests themselves.

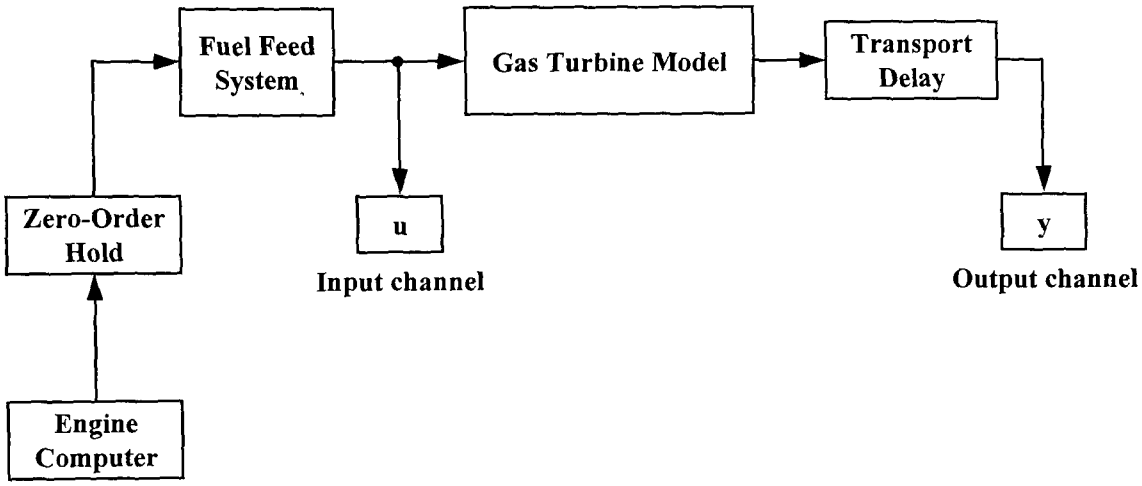

Figure 9-13. Simulated engine test set-up.

The two possible sources of error were studied separately, by conducting three different simulations: of the engine model alone, of the engine with the time delay and finally of the fuel feed and engine. Discrete OE models were then estimated between the measured input and output signals and the results for each scenario are presented in Table 9-20. A negative pole is estimated at close to the unit circle if either the delay is included or if the fuel feed system band-limits the input signal.

The position of the zero is very different for each of the sources of error, since it lies outside the unit circle if the delay is included and is close to the origin with the fuel feed. The model obtained by including the fuel feed in the simulation thus closely replicates the model obtained with the real engine data, which was listed in Table 9-15.

Comparison of the equivalent s-domain models for the error-free case, when only the engine model was simulated, and the other two simulations shows that a significant bias has also been introduced into position of the dominant pole. The problem is not as severe as that for the pole-zero pair but still represents a 20% bias in the equivalent s-domain pole positions.

Further work is clearly required to analyse in greater detail the effects on an estimated discrete model of violating the ZOH assumption. Such an analysis would need to begin with the work conducted by Schoukens *et al.* (1994a), who derived a frequency-domain expression for the error introduced. An analysis of the effect being modelled by the negative pole and the zero close to the origin would also be required, in both the time and frequency domains. This effect would then need to be related to the error introduced by the violation of the ZOH assumption. It is interesting to note that the combination of a negative pole close to the unit circle and a zero close to the origin represents a lightly damped oscillatory mode, which has little effect on the amplitude of the model frequency response but introduces a significant positive phase shift.

TABLE 9-20

DISCRETE MODELS ESTIMATED FROM SIMULATED ENGINE DATA

Scenario	Discrete Zero	s-domain equivalent	Discrete Poles	s-domain equivalents
Engine Model Only	0.9341	-0.34	0.9465	-0.60
			0.8866	-0.27
With Time Delay	-1.1480	-0.02	0.9083	-0.48
			-0.9965	—
With Fuel Feed	0.0839	-0.01	0.9097	-0.47
			-0.9975	—

9.3.4 Resampling the Data

It was noticed during the simulations that increasing the sampling frequency had a significant effect on the position of the negative pole. This effect was further investigated by resampling the original gas turbine data at a higher frequency. Since the data are periodic, the necessary interpolation can be performed using the FFT, by transforming to the frequency domain and then varying the sampling frequency of the inverse FFT. Of course, the usual restriction applies that the new sampling frequency must be an integer multiple of the signal fundamental frequency.

The HP shaft data of the 30 odd harmonic signal were resampled at progressively higher frequencies in this way, until a sampling period approaching that of the estimated pure time delay was approached. Faster sampling has the effect of reducing the error introduced by the violation of the ZOH assumption and also the error due to the delay.

The results are shown in Table 9-21, which shows that the negative pole is eliminated if the data are resampled at a frequency at least four times greater than the original rate. The positions of the discrete poles and zeros, and hence their equivalents in the s-domain, depend significantly on the sampling period. The quality of the model fit to the FRF improves as the sampling period is reduced, as seen in Figure 9-14 for the model estimated with a sampling period of 12.5 ms. The equivalent s-domain poles and zeros at this sampling frequency are very close to those of the frequency-domain model, which is also listed in Table 9-21, in order to allow a comparison.

These results show that it is possible to estimate discrete models which are equivalent to those estimated in the frequency domain but that the quality of the discrete models depends strongly on the sampling period. If the sampling period is chosen as equal to the pure time delay a good model will result, the problem is that such information is not normally known *a priori*. The sampling rate was increased to $16f_s$ (80 Hz) in order to achieve a discrete model equivalent to that estimated in the frequency domain. This is far higher than the sampling rates normally recommended for time-domain estimation, which were discussed in Chapter 3. For example, the criterion proposed by Isermann (1980) would result in a sampling frequency of between 3 Hz and 9 Hz in this case.

In addition, as the sampling rate is increased the numerical sensitivity deteriorates, due to the clustering of discrete poles and zeros around the point $z = 1$. This is not desirable if close pole-zero pairs are to be accurately identified. The use of frequency-domain techniques to directly estimate s-domain models is seen to be a far more straightforward approach for this application, where a physical interpretation is to be made of the model poles and zeros.

TABLE 9-21

DISCRETE MODELS ESTIMATED FROM RESAMPLED ENGINE DATA

Sampling Period (ms)	Delay (ms)	Discrete Zero	s-domain equivalent	Discrete Poles	s-domain equivalents
200	0	0.0225	-0.0338	-0.9936	—
(no resampling)				0.9094	-0.47
50	0	0.9771	-0.4625	0.9836	-0.3298
($4f_s$)				0.9658	-0.6968
25	0	0.9932	-0.2731	0.9942	-0.2309
($8f_s$)				0.9858	-0.5710
16.67	16.67	0.9923	-0.4610	0.9945	-0.3292
($12f_s$)				0.9885	-0.6958
12.5	12.5	0.9955	-0.3570	0.9965	-0.2823
($16f_s$)				0.9923	-0.6191
Frequency-domain	13.2	—	-0.3403	—	-0.2749
model				—	-0.6017

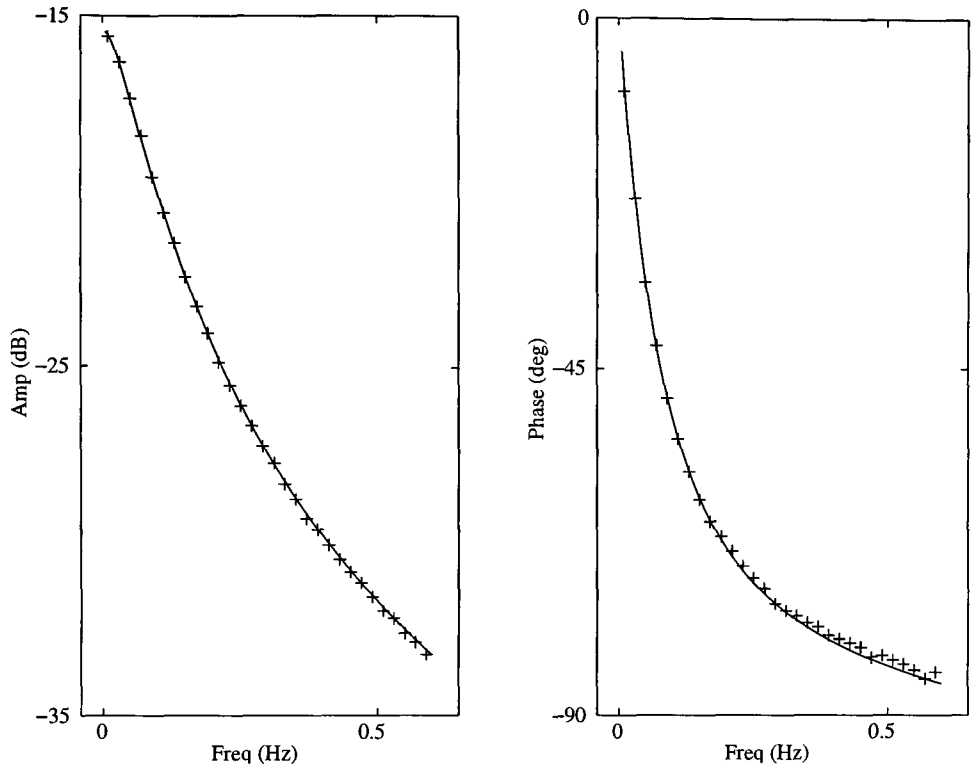

Figure 9-14. Frequency response of discrete model estimated using resampled data (solid) and FRF (crosses), for 1/2 model, HP shaft.

9.4 Conclusions

The frequency-domain identification of the fuel flow to shaft speed dynamics of a twin-shaft gas turbine has been studied, with the aim of verifying existing thermodynamic engine models. Frequency-domain methods are better suited to the aims of this investigation and to the measurement set-up employed. The s-domain models required for verifying the thermodynamic models can be directly estimated and the use of a measured, noise corrupted, input signal is assumed as part of the estimator model.

The results presented in this chapter show the potential of employing broad-band multisines as test signals. Such signals can be designed to concentrate their whole power within the system bandwidth. A sufficiently high reconstruction frequency can then be used, so that the fuel feed system filters out the reconstruction components. This overcomes the problems caused by the lack of anti-aliasing filters in the measurement set-up. The flexibility of multisine signals has been illustrated by the range of signals applied to the engine.

High quality models of each shaft were estimated using the 30 odd harmonic data and their fit to the frequency data was illustrated. The results were cross-validated on a data-set of 15 quasi-log harmonics and the estimated models found to be very similar. The pure time delay was included as a parameter for estimation and the results were close to those obtained in previous studies for a typical gas turbine.

The estimated models show the HP shaft dynamics to be first-order and the LP shaft dynamics second-order. The models thus match those derived by Fitchie *et al.* (1957) for a turbojet engine and the linearised thermodynamic models of the Spey engine, in terms of overall structure. A comparison of the models showed that the shaft dynamics appear to be faster than predicted by the thermodynamic models and that the second-order dynamics of the LP shaft are also more significant than predicted.

Additional low frequency effects were detected on both shafts which could be modelled by the addition of a close pole-zero pair. The influence of these low frequency modes was studied by excluding the lower frequencies from the data set and re-estimating a range of models. The time constants of the additional pole-zero pairs are too slow to be associated with the shaft dynamics and suggest that they are modelling thermal effects, which are not incorporated in the thermodynamic models. Further work is currently being conducted to verify the presence of these modes, using test signals with a lower fundamental frequency.

A presence of a weak even-power engine nonlinearity was detected during the data analysis presented in Chapter 8. The influence of this nonlinearity was apparent in the models estimated using the consecutive harmonic data. A study of the autocorrelation of the frequency-domain residuals confirmed the influence of the nonlinearity, which had a significant effect on the models estimated with the consecutive multisine.

It was thus possible to use frequency-domain techniques to estimate linear engine models which were free from the influence of any significant nonlinear effects. It can be stated with some confidence that the small signal dynamics of the shafts have been accurately identified, which is shown clearly by the quality of the model fits to the frequency data. The linear and nonlinear effects are clearly separated in the data; and the shaft dynamics and the possible heat transfer dynamics are clearly separated in the estimated s-domain models.

Discrete models were also estimated in the time-domain and models with good input-output properties were obtained. The problem was that the preferred models for each shaft contained a single real negative pole. Such a pole has no continuous counterpart and cannot be compared to the poles of the s-domain thermodynamic models. A range of possible causes of this effect were discussed and it was shown through simulation that it can be caused by the presence of a pure time delay shorter than the sampling period, or an actuator which band-limits the input signal and causes the ZOH assumption to be violated. Further analysis of this problem is required, as outlined at the end of Section 9.3.3.

The error introduced by both of these phenomena is reduced by faster sampling and this was investigated by resampling the engine data at progressively faster rates. The use of periodic inputs signals allows the data to be resampled in a straightforward manner using the FFT. The negative pole was eliminated in this way and a high quality discrete model was obtained when the sampling period approached the value of the pure time delay. The equivalent s-domain poles and zero of this model were very close to those estimated directly in the frequency domain.

It is concluded that it is possible to estimate discrete models which are equivalent to the frequency-domain models but that the results strongly depend on the sampling frequency selected. Estimation of discrete models remains valid if the models are required for simulation or controller design but is not the best approach for this specific problem, which is to verify the thermodynamic engine models.

This work has produced a large body of results, based on limited testing of a gas turbine. It illustrates the potential of frequency-domain techniques for modelling industrial systems where a physical interpretation of the model parameters is to be made. This is particularly true if the system contains a pure time delay or the input signal has to be measured in order to eliminate the input actuator dynamics.

Chapter X

Conclusions

Abstract — In this final chapter the major contributions and conclusions of the thesis are summarised and some ideas for further research are presented.

This thesis has dealt with a number of aspects of the application of modern system identification techniques to engineering problems. The main emphasis has been on the design of appropriate test signals to excite the systems to be modelled. Signal design has been examined in terms of the influence of both stochastic and nonlinear effects on linear system identification. The design of appropriate signals to measure points on frequency-domain Volterra kernels has also been addressed. The test signals chosen for detailed study where multisines, since they allow a completely arbitrary input spectrum to be defined. The flexibility afforded by this choice was exploited throughout the thesis.

It was stated at the outset that the thesis could be divided into two parts. The first of these was a study of test signal design, which was dealt with in Chapters 3 to 6. The second was an application example, dealing with the testing and modelling of an aircraft gas turbine, presented in Chapters 7 to 9. The main contributions and conclusions for each of these parts will now be dealt with in turn.

10.1 Test Signal Design

The present study of test signal design was motivated by the need to take full advantage of the potential represented by multisine signals, since these signals can now be easily and accurately generated using modern computer technology. The great advantage of multisines is their flexibility and this was fully exploited in the design approaches proposed in this thesis. However, if periodic signals are to be used then many of the engineering assumptions about nonlinear effects being "averaged out" by the use of random excitations are no longer valid. It was for this reason that the study of the influence of nonlinearities on multisine signals was undertaken.

In Chapter 3, the main issues relating to the design of the identification experiment were discussed, with particular emphasis being placed on the choice of input signals. The use of periodic signals was motivated and the flexibility in design afforded by multisine signals was stressed. Two measures of signal quality were considered, the widely known Crest Factor and the more recently defined Time Factor. Techniques for minimising the CF of multisines were discussed and it was shown that the CF of consecutive and odd multisines of 30 or more frequencies are equal to, or less than, that of single sines when using the $L\infty$ approach proposed by Guillaume (1992).

The relative test times for FRF measurements using single sine and multisine signals was then examined in some detail. The settling times associated with each approach were taken into account and the test times required to obtain equal accuracy of FRF estimates were then calculated. It was shown that if the multisine signals have the same CF as the single sines, the reduction in test times using multisines is derived entirely from the reduction in settling time. A further reduction can be achieved if the multisine CF is less than that of the single sines, which emphasises the importance of CF minimisation. It was also shown that in the case of single sines and multisines, the TF reduces to half the squared CF and provides no additional information about the signal quality.

The design of MLBS signals was also discussed and the problems with the time-domain design criteria commonly employed were illustrated. An alternative frequency-domain approach was proposed, based on placing the half power point of the MLBS spectrum close to the maximum frequency of interest and then adjusting the sequence length to place an appropriate number of frequencies within that range. An MLBS signal designed in this way was shown to have comparable spectral properties to a multisine with $L\infty$ phases, in the frequency range of interest, and hence comparable test times. The TF was seen to be a useful measure for discriminating between different MLBS designs.

The main contribution of Chapter 3 was to study in detail the relative test times using single sine and multisine signals. If equal accuracy is required then the benefits of multisine signals are derived almost entirely from the reduction in settling time. In practice, if the SNRs are good then considerably less averages of the multisine may be required in order to achieve an acceptable accuracy. This point was illustrated in Chapter 8 using the gas turbine data. The benefit derived from the reduction in settling time should thus be thought of as the minimum achievable using multisine signals.

A novel methodology was presented in Chapter 4 for studying the frequency-domain contributions generated by static polynomial nonlinearities, which was then extended to nonlinear systems with memory. This analysis divided the contributions generated at the input frequencies by odd-order nonlinear terms into two types. The Type I contributions cannot be influenced by the selection of the input harmonics and depend only on the order of the nonlinearity and on the number and amplitude of the harmonics at the input to the nonlinear element. By contrast, the number of Type II contributions which fall at the input frequencies depends entirely on the specific harmonics included in the signal.

The new analysis method provides a far better insight into the influence of the nonlinearity than the conventional approach, which is to simply divide the nonlinear contributions into *n*th harmonics and intermodulation harmonics. This was illustrated by the test signals subsequently designed in the chapter, based on the new analysis approach. These signals were termed *no interharmonic distortion* (NID) multisines since they do not generate any Type II contributions at the input frequencies, for a given maximum order of nonlinearity.

It was shown that NID signals can be used to measure the underlying linear dynamics of a Wiener-Hammerstein nonlinear system, multiplied by a gain factor. This gain factor can be considered as a constant a value at each excited frequency, if the number of frequencies is large, and depends on the harmonic amplitudes of the input signal and the dynamics of the first linear system.

This was shown to be closely related to work conducted by Schoukens *et al.* (1998) using multisines with uniform random phases, averaged over several hundred experiments. The system measured in this way is defined as the *related linear dynamic system* and can be considered as the best linear approximation of the overall nonlinear system. This approximation is dependent on the total power of the input signal and a given result will be specific to that signal and the input amplitude used. Application of a NID signal removes the need to average across many experiments, with the obvious restriction that a multisine with only the specified harmonics must be imposed on the system.

Chapter 5 examined the influence of an odd-order nonlinearity on a range of multisine signals, by analysing the interaction of Type I and Type II contributions at the test frequencies. The aim in this case was to isolate the linear component of the system response and not, as in Chapter 4, to obtain the best linear approximation of the overall nonlinear system. The linear dynamics identified in this way will be independent of the input signal and its amplitude.

This work was motivated by the need to understand how nonlinear distortions will affect linear system identification and to design signals which are, as far as possible, robust to nonlinear effects. The influence of odd-order nonlinearities was studied since the influence of even-order nonlinear terms can be easily eliminated by simply omitting all the even harmonics from the input signal.

An important conclusion was reached that minimising the number of nonlinear contributions which fall at a given test frequency does not minimise the nonlinear distortion. There is consequently no benefit in using very sparse signals, with prime or NID harmonics, if the aim is to reduce the nonlinear distortion at the test frequencies to the minimum possible. The key factor is the signal CF and an odd harmonic multisine is therefore recommended for general testing, since it is immune to even-order nonlinearities and amenable to CF minimisation.

Since a nonlinear error is inherent with any odd-order nonlinearity a compensation technique must be employed if it is to be completely eliminated. A method based on testing the system at two signal amplitudes was extended to the multisine case and practical results were presented which illustrated its application. The advantage of this approach is that it can be applied without the need to specify a particular nonlinear model. The influence of the nonlinearity can be quantified by calculating the complex error between the uncompensated and compensated data.

An overall identification scheme was proposed at the end of Chapter 5, for identifying linear models in the presence of nonlinear distortions. The steps involved include: an initial pilot test using a sparse odd signal, further testing using an odd harmonic signal, elimination of the nonlinear effects using a two level technique and, finally, the estimation of a linear model. The systematic application of this approach will ensure that high quality linear models are identified, from which the nonlinear influence has been excluded. Alternatively, the identification scheme may indicate that a linear model is inappropriate for the system under test.

It is clear from the work presented in Chapters 4 and 5 that the signals required to address the aims of each chapter are very different. If the aim is to control the nonlinear influence at the test frequencies, in order to measure the best linear approximation, then sparse odd multisines are required. If the aim is to minimise the nonlinear influence, in order to measure the underlying linear system, then low CF odd multisines with dense spectra should be used.

Chapter 6 addressed the specific problem of designing signals to directly measure points on the frequency-domain Volterra kernel of a nonlinear system. A study of previously designed signals of this type highlighted many of their drawbacks. The new designs overcome a number of these problems, by allowing the maximum possible kernel points to be measured and maintaining a near-even harmonic spacing. Signals with paired

harmonics can be used to measure points close to the unmeasured kernel diagonals. The main drawback associated with all such signals is that the test times are considerably longer than those for linear systems with the same bandwidth, which suggests that they are only suitable for testing systems with fast dynamics. However, these periodic signals are an attractive alternative to the use of random signals for the measurement of Volterra kernels. Their practical application was illustrated by measuring the second-order kernel of a dc servo motor system.

10.2 Gas Turbine Testing and Modelling

The frequency-domain identification of the fuel flow to shaft speed dynamics of a twin-shaft aircraft gas turbine was studied in the second part of the thesis. The main aim of this work was to verify the linearised thermodynamic models derived from the engine physics. It was shown that frequency-domain methods are better suited to the aims of this investigation and to the measurement set-up employed. The s-domain models required for verifying the thermodynamic models can be directly estimated and the use of a measured, noise corrupted, input signal is assumed as part of the estimator model. The application of multifrequency test signals has been motivated by the need to reduce engine test times, in order to make systematic engine testing a cost-effective option. These results show the potential of employing broad-band multisines as test signals.

A detailed overview of work previously conducted in this area was presented in Chapter 7, which showed that, while a great deal of work has been conducted on gas turbine modelling, the systematic application of modern system identification techniques to this problem is still at an early stage. The system identification work conducted to date has concentrated on the time-domain estimation of discrete models. Such models are extremely useful for purposes of engine simulation and control system design but they cannot be interpreted in terms of physical parameters if they are estimated using a band-limited input signal, due to the violation of the ZOH assumption.

The specific aims of this investigation were thus defined as:

- Applying wide-band, low CF, multisine signals to the testing of gas turbines.
- Estimating many points on the engine FRF in one test, along with a measure of the uncertainty on the estimates.

- Directly estimating *s*-domain models using frequency-domain techniques, which could then be compared to the linearised *s*-domain thermodynamic models.
- Systematically assessing the influence of system nonlinearities for the selected input amplitude range and their influence on the estimated linear models.
- Estimating the engine pure time delay as part of the *s*-domain model.

Particular emphasis was placed on identifying the linear second-order effects, which are the result of shaft interactions. The possible influence of thermal effects, which are not incorporated into the thermodynamic models, was also investigated.

The design of a range of test signals for gas turbine testing was described in detail in Chapter 8, based on *a priori* knowledge of the engine dynamics. A particular feature of the gas turbine test set-up is that the break-point of the slowest fuel-feed pole is only six times higher than that of the fastest pole of the engine shaft dynamics. This means that the fuel feed acts to band-limit the input signal before it is measured.

The new signal designs were successfully applied to the engine and a detailed analysis of the results was presented. It was found that the data gathered from the engine showed good clock synchronisation and low drift and good SNRs were obtained, of 40 dB or better. The influence of an even-order nonlinearity was detected in the engine, which would not affect the odd harmonic multisines used. The FRFs of the HP and LP shafts were estimated for a range of tests and the uncertainty of the estimates was found to be very small, for input amplitudes of $\pm$ 5% W_f or greater. Thus an acceptable accuracy was achieved after averaging across only six periods of the multisine signal. This shows that broad-band multisine signals can be applied with confidence to engine testing, in place of the more commonly used single sine tests, leading to much reduced test times and hence reduced costs.

High quality engine models were then estimated in Chapter 9, with the pure time delay included as an estimated parameter. The models showed the HP *shaft dynamics* to be first-order and the LP *shaft dynamics* to be second-order. The estimated values of pure time delay were close to those obtained in previous studies for a typical gas turbine. A comparison of the model dynamics with those of the thermodynamic models showed that the shaft dynamics appear to be faster than previously predicted and that the second-order dynamics of the LP shaft are also more significant.

Additional low frequency effects were detected on both shafts which could be modelled by the addition of close pole-zero pairs. The time constants of the additional pole-zero pairs are too slow to be associated with the shaft dynamics and suggest that they are modelling thermal effects, which are not incorporated in the thermodynamic models.

It was thus possible to use frequency-domain techniques to estimate linear engine models, with low uncertainty, which were also free from the influence of any significant nonlinear effects. It can be stated with some confidence that the small signal dynamics of the shafts have been accurately identified, which is shown by the quality of the model fits to the frequency data. The linear and nonlinear effects are clearly separated in the test data and, in addition, the shaft dynamics and the possible heat transfer dynamics are clearly separated in the estimated s-domain models.

Discrete models were also estimated for comparison, using a time-domain approach, and models with good input-output properties were obtained. The problem was that the preferred models for each shaft contained a single real negative pole. Such a pole has no continuous counterpart and cannot be compared to the poles of the s-domain thermodynamic models. A range of possible causes of this effect were studied and it was shown through simulation that it can be caused by the presence of a pure time delay shorter than the sampling period or an actuator which band-limits the input signal, hence violating the ZOH assumption.

The error introduced by both of these phenomena is reduced by faster sampling and this was investigated by resampling the engine data at progressively faster rates. The use of periodic inputs signals allows the data to be resampled without any approximation. The negative pole was eliminated in this way and a high quality discrete model was obtained when the sampling period approached the value of the pure time delay. The equivalent s-domain poles and zero of this model were very close to those estimated directly in the frequency domain.

It was concluded that it is possible to estimate discrete models which are equivalent to the frequency-domain models but that the results strongly depend on the sampling frequency selected. Estimation of discrete models remains valid if the models are required for simulation or controller design but is not the best approach for this specific problem, that of verifying the thermodynamic engine models.

This work has produced a large body of results, based on limited testing of a gas turbine, and each of the aims of the investigation has been achieved. This illustrates the potential of frequency-domain techniques for modelling industrial systems where a physical interpretation of the model parameters is to be made. This is particularly true if the system contains a pure time delay or the input signal has to be measured in order to eliminate the input actuator dynamics.

10.3 Ideas for Further Research

The following are possible areas for further work on test signal design:

- Development of a frequency-domain CF minimisation technique. The relationship between the CF of a multisine signal and the way that Type I and Type II nonlinear contributions interact at the signal frequencies was discussed in Chapter 5. It is clear that low CF signals generate Type I contributions which act in anti-phase to the Type II contributions. This observation could form the basis of a frequency-domain approach to CF minimisation, whereby the signal phases are selected to maximise this effect.
- Use of sparse odd multisines under non-ideal conditions. The properties of both NID signals and Volterra measuring signals depend on exciting the system with a precise input spectrum. This may not always be possible due to the presence of input noise, nonlinearities in the signal generation equipment or loading effects, which generate power at unwanted frequencies. It would be worthwhile to investigate the robustness of the signal properties in the face of such input disturbances.
- The identification of the cascaded linear subsystems of a Wiener-Hammerstein model, discussed in Chapter 4, and the measurement of Volterra kernels, described in Chapter 6, were based on the assumption that the systems under test could be described by such models. There is a need to develop straightforward structure detection methods for block-oriented nonlinear models such as the Wiener-Hammerstein structure and also for establishing if the application of Volterra models is appropriate.
- The complexity associated with the measurement and validation of the nonparametric Volterra kernels also needs to be addressed. This is particularly true for higher-order kernels, where the "curse of dimensionality" comes into play and a large number of points are required to adequately characterise a

nonparametric kernel. The associated difficulties could be greatly reduced by estimating parametric Volterra kernels, using the measured frequency points as data.

There is also considerable potential for further work in the area of gas turbine modelling, namely:

- Validation of the results presented in this thesis through further testing and testing at other operating points to allow the verification of the linearised thermodynamic models across the complete engine operating range.
- Testing with signals with lower fundamental frequencies to verify the presence of the weak low frequency modes detected in the current study.
- Use of signals with larger input amplitudes for nonlinear detection and modelling, in order to excite the engine nonlinearities and allow them to be modelled.
- Direct estimation of *s*-domain state space models, which would correspond directly with the structure of the linearised thermodynamic models.
- An analytical study of the conditions under which single real negative poles are estimated in discrete-time models, in order to more firmly establish the error, or errors, which cause this effect.

References

Andersson, T., P. Pucar and L. Ljung (1994). "Identification aspects of inter-sample input behaviour", in *Preprints 10th IFAC Symp. on System Identification*, Copenhagen, vol. 3, pp. 137-142.

Åström, K.J. (1980). "Maximum likelihood and prediction error methods", *Automatica*, vol. 16, pp. 551-574.

Åström, K.J., P. Hagander and J. Sternby (1984). "Zeros of sampled systems", *Automatica*, vol. 20, pp. 31-38.

Barker, H.A. and R.W. Davy (1975). "System identification using pseudorandom signals and the discrete Fourier transform', *Proc. IEE*, vol. 122, no. 3, pp. 305-311.

Barker, H.A. and W. Davy (1978). "Measurement of second-order Volterra kernels using pseudorandom ternary signals", *Int. J. of Control*, vol. 27, no. 2, pp. 277-291.

Barker, H.A. and M.H. Al-Hilal (1985). "Nonlinear system identification using pseudorandom signals with partially orthogonal transforms", in *Proc. 7th IFAC Symp. on Identification and System Parameter Estimation*, York, pp. 415-420.

Barker, H.A. (1993). "Design of multi-level pseudo-random signals for system identification", chapter 11 of *Perturbation Signals for System Identification*, K. Godfrey (ed.), Prentice Hall, New York.

Bauerfeind, K. (1968). "A new method for the determination of transient jet engine performance based on the non-stationary characteristics of the components", *AGARD Conf. Proc. No. 34 - Advanced Components for Turbojet Engines*, part 2, paper 32, pp. 1-23.

Bauerfeind, K. (1982). "Some general topics in the field of engine handling", *AGARD Conf. Proc. No. 324 - Engine Handling*, paper 1, pp. 1-14.

Bendat, J.S. and A.G. Piersol (1980). *Engineering Applications of Correlation and Spectral Analysis*, Wiley-Interscience, New York.

Bendat, J.S. (1990). *Nonlinear System Analysis and Identification from Random Data*, Wiley Interscience, New York.

Billings, S.A. and S.Y. Fakhouri (1979). "Identification of systems composed of linear dynamic and static nonlinear elements", in *Proc. 5th IFAC Symp. on Identification and System Parameter Estimation*, Darmstadt, pp. 493-500.

Billings, S.A. (1980). "Identification of nonlinear systems - a survey", *IEE Proc. Control Theory and Applications*, vol. 127, no. 6, pp. 272-285.

Billings, S.A. and K.M. Tsang (1989). "Spectral analysis for non-linear systems, part I: parametric non-linear spectral analysis", *J. Mechanical Systems and Signal Processing*, vol. 3, no. 4, pp. 319-339.

Billings, S.A. and K.M. Tsang (1990). "Spectral analysis of block-structured nonlinear systems", *J. Mechanical Systems and Signal Processing*, vol. 4, no. 2, pp. 117-130.

Boyd, S., Y.S. Tang and L.O. Chua (1983). "Measuring Volterra kernels", *IEEE Trans. Circuits and Systems*, vol. 30, no. 8, pp. 571-577.

Boyd, S. and L.O. Chua (1983). "Uniqueness of a basic nonlinear structure", *IEEE Trans. Circuits and Systems*, vol. 30, no. 9, pp. 648-651.

Boyd, S., L.O. Chua and C.A. Desoer (1984). "Analytical foundations of Volterra series", *IMA J. Mathematical Control and Information*, vol. 1, pp. 243-282.

Boyd, S. (1986). "Multitone signals with low crest factor", *IEEE Trans. Circuits and Systems*, vol. 33, no. 10, pp. 1018-1022.

Brillinger, D.R. (1975). *Time Series Data Analysis and Theory*, Holt, Reinhart and Winston, New York.

Bussgang, J.J., L. Ehrman and J.W. Graham (1974). "Analysis of nonlinear systems with multiple inputs", *Proc. IEEE*, vol. 62, pp. 1088-1119.

Chen, X. and T.W. Parks (1987). "Design of FIR filters in the complex domain", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 35, no. 2, pp. 144-153.

Chua, L.O. and C.-Y. Ng (1979). "Frequency domain analysis of nonlinear systems: general theory", *Electronic Circuits and Systems*, vol. 3, no. 4, pp. 165-185.

Chua, L.O. and Y. Liao (1989). " Measuring Volterra Kernels (II)", *Int. J. Circuit Theory and Applications*, vol. 17, pp. 151-190.

Chua, L.O. and Y. Liao (1991). " Measuring Volterra kernels III: how to estimate the highest significant order", *Int. J. Circuit Theory and Applications*, vol. 19, pp. 189-209.

Cohen, H., G.F.C. Rogers and H.I.H. Saravanamuttoo (1987). *Gas Turbine Theory*, Longman Scientific, New York.

Cottington, R.V. (1974). "Total powerplant simulation", *AGARD Conf. Proc. No. 151 - Power plant controls for aero-engine turbines*, paper 24, pp. 1-D2.

Cottington, R.V. and C.B. Pease (1979). "Dynamic response testing of gas turbines", *ASME J. Engineering for Power*, vol. 101, pp. 95-100.

Dadd, G.J., A.E. Sutton and A.W.M. Greig (1996). "Multivariable control of military engines", *AGARD Conf. Proc. No. 572 - Advanced Aero-Engine Concepts and Controls*, paper 28, pp. 1-12.

Duym, S. and J. Schoukens (1995). "Design of excitation signals for the restoring force surface method", *J. Mechanical Systems and Signal Processing*, vol. 9, no. 2, pp. 139-158.

Evans, C., D. Rees and L. Jones (1992). "Design of test signals for identification of linear systems with nonlinear distortions", *IEEE Trans. Instrumentation and Measurement*, vol. 41, no. 6, pp. 768-774.

Evans, C., D. Rees and L. Jones (1994a). "Nonlinear disturbance errors in system identification using multisine test signals", *IEEE Trans. Instrumentation and Measurement*, vol. 43, no. 2, pp. 238-244.

Evans, C., D. Rees, L. Jones and D. Hill (1994b). "Measurement and identification of gas turbine dynamics in the presence of noise and nonlinearities", in *Proc. IEEE Instrumentation and Measurement Technology Conf.*, Japan, vol. 2, pp. 609-614.

Evans, C., D. Rees, L. Jones and D. Hill (1994c). "Time and frequency domain identification of jet engine dynamics: problems and solutions", in *Preprints 10th IFAC Symp. on System Identification*, Copenhagen, vol. 2, pp. 243-248.

Evans, C., D. Rees and L. Jones (1995). "Identifying linear models of systems suffering nonlinear distortions, with a gas turbine application", *IEE Proc. Control Theory and Applications*, vol. 142, no. 3, pp. 229-240.

Evans, C., D. Rees, L. Jones and M. Weiss (1996). "Periodic signals for measuring nonlinear Volterra kernels", *IEEE Trans. Instrumentation and Measurement*, vol. 45, no. 2, pp. 362-371.

Evans, C., D. Rees and D. Hill (1998). "Frequency-domain identification of gas turbine dynamics". Accepted for the *IEEE Trans. Control Systems Technology*, scheduled for publication in vol. 6, no. 5, September 1998.

Eykhoff, P. (1974). *System Identification: Parameter and State Estimation*, Wiley Interscience, Chichester.

Eykhoff, P. (1981). Prologue to *Trends and Progress in System Identification*, Pergamon Press, Oxford.

Eykhoff, P. (1984). "Identification theory: practical implications and limitations", *Measurement*, vol. 2, no. 2, pp. 75-85.

Fasol, K.H. and H.P. Jörgl (1980). "Principles of model building and identification", *Automatica*, vol. 16, pp. 505-518.

Fawke, A.J., H.I.H. Saravanamuttoo and M. Holmes (1972). "Experimental verification of a digital computer simulation method for predicting gas turbine dynamic behaviour", *Proc. Institute of Mechanical Engineers*, vol. 186, part 27, pp. 323-329.

Fawke, A.J. and H.I.H. Saravanamuttoo (1973). "Digital computer simulation of the dynamic response of a twin-spool turbofan with mixed exhausts", *UK Aeronautical J.*, vol. 77, part 753, pp. 471-8.

Fitchie, J.W., S.W. White, R.H. Cronshaw and H.J. Willcocks (1959). "An experimental and theoretical study of the response of a two shaft turbojet subjected to small disturbances", *National Gas Turbine Establishment (UK)*, Note NT 420.

Flower, J.O. and S.C. Forge (1981). "Developments in frequency-response determination using Schroeder-phased harmonic signals", *The Radio and Electronic Engineer*, vol. 51, no. 5, pp. 226-232.

Franklin, G.F., J.D. Powell and M.L. Workman (1990). *Digital Control of Dynamic Systems*, Addison-Wesley, Reading, Massachusetts.

Gade, S. and H. Herlufsen (1987). "Use of weighting functions in DFT/FFT analysis", Parts I and II, *Brüel and Kjær Technical Review*, no. 3 and no. 4.

Gifford, S.J. and G.R. Tomlinson (1989). "Recent advances in the application of functional series to non-linear structures", *J. Sound and Vibration*, vol. 135, no. 2, pp. 289-317.

Godfrey, K.R. and P.A.N. Briggs (1972). "Identification of processes with direction dependent dynamic responses", *Proc. IEE*, vol. 119, no. 12, pp. 1733-1739.

Godfrey, K.R. and D.J. Moore (1974). "Identification of processes having direction dependent responses, with gas-turbine engine applications", *Automatica*, vol. 10, pp. 469-481.

Godfrey, K.R. (1980). "Correlation methods", *Automatica*, vol. 16, pp. 527-534.

Godfrey, K.R. (1993a). "Introduction to perturbation signals for time-domain system identification", chapter 1 of *Perturbation Signals for System Identification*, K. Godfrey (ed.), Prentice Hall, New York.

Godfrey, K.R. (1993b). "Introduction to perturbation signals for frequency-domain system identification", chapter 2 of *Perturbation Signals for System Identification*, K. Godfrey (ed.), Prentice Hall, New York.

Gold, H. and S. Rosenzweig (1952). "A method for estimating speed response of gas-turbine engines", *National Advisory Committee for Aeronautics (Washington)*, Research Memorandum E51 K21.

Goodwin, G.C. and R.L. Payne (1976). "Choice of sampling interval", in *System Identification - Advances and Case Studies*, R.K. Mehra and D.G. Lainiotis (eds.), Academic Press, New York.

Goodwin, G.C. and R.L Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis*, Academic Press, New York.

Greblicki, W. and M. Pawlak (1986). "Identification of discrete Hammerstein systems using kernel regression estimates", *IEEE Trans. Automatic Control*, vol. 31, no. 1, pp. 74-77.

Guillaume, P., J. Schoukens and R. Pintelon (1989). "Sensitivity of roots to errors in the coefficients of polynomials obtained by frequency-domain estimation methods", *IEEE Trans. Instrumentation and Measurement*, vol. 38, no. 6, pp 1050-1056.

Guillaume, P., J. Schoukens, R. Pintelon and I. Kollár (1991). "Crest factor minimisation using nonlinear Chebyshev approximation methods", *IEEE Trans. Instrumentation and Measurement*, vol. 40, no. 6, pp. 982-989.

Guillaume, P. (1992). "Identification of multi-input multi-output systems using frequency-domain methods", Ph.D. dissertation, Vrije Universiteit Brussel, Department ELEC, Belgium.

Haber, R. (1985). "Nonlinearity tests for dynamic processes", in *Proc. 7th IFAC Symp. on Identification and System Parameter Estimation*, York, pp. 409-414.

Harman, R.T.C. (1981). *Gas Turbine Engineering*, Macmillan Press, London.

Hennecke, D.K. and K. Trappmann (1982). "Turbine tip clearance control in gas turbine engines", *AGARD Conf. Proc. No. 324 - Engine Handling*, paper 16, pp. 1-12.

Hermsmeyer, S. (1996). "Improved methods for modelling turbine engine gas flow properties in real time", Ph.D. dissertation, University of Birmingham, School of Electronic and Electrical Engineering.

Hill, D.C. (1994). "System identification of gas turbine engines", Ph.D. dissertation, University of Birmingham, School of Electronic and Electrical Engineering.

Hill, D.C. (1997). "Identification of gas turbine dynamics: time-domain estimation problems", *ASME Gas Turbine Conf.*, paper 97-GT-31, pp. 1-7.

Hollkamp, J.J. and S.M. Batill (1992). "Structural identification using order overspecified time-series models", *ASME J. Dynamic Systems*, Measurement and Control, vol. 114, no. 1, pp. 27-33.

Hung, G. and L. Stark (1977). "The kernel identification method (1910-1977) - review of theory, calculation, application and interpretation", *Mathematical Biosciences*, vol. 37, no. 3/4, pp. 135-190.

Isermann, R. (1980). "Practical aspects of process identification", *Automatica*, vol. 16, pp. 575-587.

Jackson, D. (1988). "Investigation of state space architectures for engine models", *Rolls Royce plc*, Report TDR 9331.

Kollár, I. (1993). "On frequency-domain identification of linear systems", *IEEE Trans. Instrumentation and Measurement*, vol. 42, no. 1, pp 2-6.

Kollár, I. (1994). *Frequency Domain System Identification Toolbox for use with Matlab*, The Mathworks Inc., Natick, Massachusetts.

Kollár, I., R. Pintelon and J. Schoukens (1994). "Frequency domain system identification toolbox for Matlab: a complex application example", in *Preprints 10th IFAC Symp. on System Identification*, Denmark, vol. 4, pp. 23-28.

Kollár, I., G. Franklin and R. Pintelon (1996). "On the equivalence of z-domain and s-domain models in system identification", in *Proc. IEEE Instrumentation and Measurement Conf.*, Brussels, pp. 14-19.

Korenberg, M.J. (1991). "Parallel cascade identification and kernel estimation for nonlinear systems", *Annals of Biomedical Engineering*, vol. 19, pp. 429-455.

Lawrence, J.O.N. and R.D. Powell (1957). "The application of servo-mechanism analysis to fuel control problems", *Proc. Institute of Mechanical Engineers*, vol. 172, pp. 439-469.

Lawrence, P.J. (1981). "Estimation of the Volterra functional series of a nonlinear system using frequency response data", *IEE Proc. Control Theory and Applications*, vol. 128, no. 5, pp. 206-210.

Lee, Y.N. and M. Schetzen (1965). "Measurement of the kernels of a nonlinear system by cross-correlation", *Int. J. Control*, vol. 2, pp. 237-254.

Ljung, L. (1987). *System Identification - Theory for the User*, Prentice Hall, Englewood Cliffs.

Ljung, L. (1994). Private correspondence with the author, Linköping, January 18th.

Ljung, L. (1995). *System Identification Toolbox for use with Matlab*, The Mathworks Inc., Natick, Massachusetts.

Lutchen, K.R., K. Yang, D.W. Kaczka and B. Suki (1993). "Optimal ventilation waveforms for estimating low-frequency respiratory impedance", *J. Applied Physiology*, vol. 74, no. 7, pp. 478-488.

Maccallum, N.R.L. (1979). "Thermal influences in gas turbine transients - effects of changes in compressor characteristics", *ASME Gas Turbine Conf.*, paper 79-GT-143, pp. 1-8.

Maccallum, N.R.L. (1981). "Further studies of the influence of thermal effects on the predicted acceleration of gas turbines", *ASME Gas Turbine Conf.*, paper 81-GT-21, pp. 1-8.

Maccallum, N.R.L. and O.F. Qi (1989). "The transient behaviour of aircraft gas turbines", presented at *IMechE Seminar on Gas Turbines - Technology and Development*, London, pp. 1-14.

Marmarelis, P.Z. and V.Z. Marmarelis (1978). *Analysis of Physiological Systems - the White Noise Approach*, Plenum Press.

Mats, E.B. and A.P. Tunakov (1982). "Requirements on modern mathematical models of gas turbine engines, II", *Soviet Aeronautica*, vol. 25, part 1, pp. 70-73.

McCormack, A.S., K.R. Godfrey and J.O. Flower (1995). "The design of multilevel multiharmonic signals for system identification", *IEE Proc. Control Theory and Applications*, vol. 142, no. 3, pp. 247-252.

Mehra, R.K. (1981). "Choice of input signals", in *Trends and Progress in System Identification*, Pergamon Press, Oxford.

Milne, R. and L. Travé-Massuyès (1994). "Real-time model based diagnosis of gas turbines", presented at *IMechE Symp. on Quantitative and Qualitative Methods for Fault Diagnosis in Process Control*, London.

Moore, D.J. (1970). "Error correction applied to dynamic analysis", *Rolls Royce Final Report*, Brochure 5B/5B, EER-5033-70, Bristol Engine Division.

Mullins, B. (1951). "Studies on the spontaneous ignition of fuels injected into a hot air stream", *National Gas Turbine Establishment (UK)*, Report R90.

Narayanan, S. (1970). "Application of Volterra series to intermodulation distortion analysis of transistor feedback amplifiers", *IEEE Trans. Circuit Theory*, vol. 17, no. 4, pp. 518-527.

Narendra, K.S. and P.G. Gallman (1966). "An iterative method for the identification of nonlinear systems using a Hammerstein model", *IEEE Trans. Automatic Control*, vol. 11, pp. 546-550.

Norton, J.P. (1986). *An Introduction to Identification*, Academic Press, London.

Ohain, H. von (1996). Foreword to *Elements of Jet Turbine Propulsion*, J.D. Mattingly, McGraw-Hill, New York.

Onions, R.A. and A.M. Foss (1982). "Improvements in the dynamic simulation of gas turbines", *AGARD Conf. Proc. No. 324 - Engine Handling*, paper 27, pp. 1-16.

Paehlike, K.-D. and H. Rake (1979). "Binary multifrequency signals - synthesis and application", in *Proc. 5th IFAC Symp. on Identification and System Parameter Estimation*, Darmstadt, vol. 1, pp. 589-596.

Palm, G. (1978). "On representation and approximation of nonlinear systems", *Biological Cybernetics*, vol. 31, pp. 119-124.

Pilidis, P. and N.R.L. Maccallum (1982). "Models for predicting tip clearance changes in gas turbines", *AGARD Conf. Proc. No. 324 - Engine Handling*, paper 17, pp. 1-10.

Pilidis, P. and N.R.L. Maccallum (1986). "The effect of heat transfer on gas turbine transients", *ASME Gas Turbine Conf.*, paper 86-GT-275, pp. 1-10.

Pintelon, R., P. Guillaume, Y. Rolain and F. Verbeyst (1992). "Identification of linear systems captured in a feedback loop", *IEEE Trans. Instrumentation and Measurement*, vol. 41, no. 6, pp 747-754.

Pintelon, R., P. Guillaume, Y. Rolain, J. Schoukens, H. Van Hamme (1994). "Parametric identification of transfer functions in the frequency domain", *IEEE Trans. Automatic Control*, vol. 39, no. 11, pp. 2245-2260.

Pollitt, J. (1982). "Keynote address", *AGARD Conf. Proc. No. 324 - Engine Handling*, pp. K1-K3.

Ream, N. (1970). "Nonlinear identification using inverse-repeat m-sequences", *Proc. IEE*, vol. 117, no. 1, pp. 213-218.

Rees, D. (1976). "Digital processing of system responses", Ph.D. dissertation, The Polytechnic of Wales, Faculty of Engineering, Department of Electrical Engineering.

Rees, D. (1986). "System identification using composite frequency signals with low peak factors", in *Proc. Int. AMSE Conf. on Modelling and Simulation*, Sorrento, vol. 1.1 & 2, pp. 13-24.

Rees, D. and D.L. Jones (1991). "Design and application of non-binary low peak factor signals for system dynamic measurement", in *Proc. IEE Control '91 Conf.*, Edinburgh, vol. 1, pp. 644-650.

Rolt, A. (1993). "LP spool accessory drives", presented at *IMechE Seminar on The Future of Engine Accessory Drives*, Birmingham, pp. 1-12.

Saravanamuttoo, H.I.H. and A.J. Fawke (1970). "Simulation of gas turbine dynamic performance", *ASME Gas Turbine Conf.*, paper 70-GT-23, pp. 1-8.

Saravanamuttoo, H.I.H. and B.D. MacIsaac (1982). "An overview of engine dynamic response and mathematical modeling concepts", *AGARD Conf. Proc. No. 324 - Engine Handling*, paper 25, pp. 1-13.

Saravanamuttoo, H.I.H. (1992). "Overview on basis and use of performance prediction methods ", *AGARD Lecture Series No. 183 - Steady and Transient Performance Prediction of Gas Turbine Engines*, paper 1, pp. 1-18.

Schetzen, M. (1980). *The Volterra and Wiener Theories of Nonlinear Systems*, Wiley Interscience, New York.

Schoukens, J. and J. Renneboog (1986). "Modeling the noise influence on the Fourier coefficients after a discrete Fourier transform", *IEEE Trans. Instrumentation and Measurement*, vol. 35, no. 3, pp. 278-286.

Schoukens, J., R. Pintelon and J. Renneboog (1988). "A maximum likelihood estimator for linear and nonlinear systems - a practical application of estimation techniques in measurement problems", *IEEE Trans. Instrumentation and Measurement*, vol. 37, no. 1, pp. 10-17.

Schoukens, J. and R. Pintelon (1991). *Identification of Linear Systems*, Pergamon Press, Oxford.

Schoukens, J., P. Guillaume and R. Pintelon (1993a). "Design of broadband excitation signals", chapter 3 of *Perturbation Signals for System Identification*, K. Godfrey (ed.), Prentice Hall, New York.

Schoukens, J., Y. Rolain, L. Montecelli and C. De Locht (1993b). "Identification of linear systems in the presence of non-linear distortions", in *Proc. 11th Int. Modal Analysis Conf.*, Florida, vol. 1, pp. 479-485.

Schoukens, J., R. Pintelon and H. Van Hamme (1994a). "Identification of linear dynamic systems using piecewise constant excitations: use, misuse and alternatives", *Automatica*, vol. 30, no. 7, pp 1153-1169.

Schoukens, J., R. Pintelon and P. Guillaume (1994b). "On the advantages of periodic excitation in system identification", in *Preprints 10th IFAC Symp. on System Identification*, Copenhagen, vol. 3, pp. 153-158.

Schoukens, J., R. Pintelon, G. Vandersteen and P. Guillaume (1997). "Frequency-domain system identification using non-parametric noise models estimated from a small number of data sets", *Automatica*, vol. 33, no. 6, pp. 1073-1086.

Schoukens, J., T. Dobrowiecki and R. Pintelon (1998). "Parametric and nonparametric identification of linear systems in the presence of nonlinear distortions - a frequency domain approach", *IEEE Trans. Automatic Control*, vol. 43, no. 2, pp. 176-190.

Schroeder, M.R. (1970). "Synthesis of low peak-factor signals and binary sequences of low auto-correlation", *IEEE Trans. Information Theory*, vol. 16, pp 85-89.

Shanmugam, K.S. and M. Lal (1976). "Analysis and synthesis of a class of nonlinear systems", *IEEE Trans. Circuits and Systems*, vol. 23, no. 1, pp. 17-25.

Söderström, T. (1975). "Test of pole-zero cancellation in estimated models", *Automatica*, vol. 11, no. 7, pp. 537-539.

Staff at Section APD5 (1993). "Thermodynamic model of the Spey engine. Linearised for Rolls Royce by Stirling Dynamics Ltd", *DRA Pyestock (UK)*.

Stremler, F.G. (1982). *Introduction to Communications Systems*, Addison-Wesley, Reading, Massachusetts.

Suki, B. and K.R. Lutchen (1992). "Pseudorandom signals to estimate apparent transfer and coherence functions of nonlinear systems: applications to respiratory mechanics", *IEEE Trans. Biomedical Engineering*, vol. 39, no 11, pp. 1142-1151.

Thomson, B. (1974). "Basic transient effects of aero gas turbines", *AGARD Conf. Proc. No. 151 - Power Plant Controls for Aero-Gas Turbine Engines*, paper 2, pp. 1-16.

Unsigned article, (1994). "Aero power comes down to earth", *IEE Review*, January, p. 2.

Van den Bos, A. (1967). "Construction of binary multifrequency test signals", in *Proc. 1st IFAC Symp. on Identification in Automatic Control Systems*, Prague, paper 4.6.

Van den Bos, A. (1970). "Estimation of parameters of linear system using periodic test signals", Ph.D. dissertation, Technische Hogeschool Delft, The Netherlands.

Van den Bos, A. and R.G. Krol (1979). "Synthesis of discrete-interval binary signals with specified Fourier amplitude spectra", *Int. J. Control*, vol. 30, no. 5, pp. 871-884.

Van den Bos, A. (1987). "A new method for synthesis of low-peak factor signals", *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 35, no. 1, pp. 120-122.

Van den Bos, A. (1993). "Periodic test signals - properties and use", chapter 4 of *Perturbation Signals for System Identification*, K. Godfrey (ed.), Prentice Hall, New York.

Van den Eijnde, E. and J. Schoukens (1991). "On the design of optimal excitation signals", in *Preprints 9th IFAC Symp. on System Identification*, Budapest, pp. 827-832.

Van der Ouderaa, E. and J. Schoukens and J. Renneboog (1987). "Comments on 'Multitone signals with low crest factor'", *IEEE Trans. Circuits and Systems*, vol. 34, no. 9, pp. 1125-1127.

Van der Ouderaa, E. (1988). "Design of optimal input signals with minimal crest factor", PhD. dissertation, Vrije Universiteit Brussel, Department ELEC, Belgium.

Van der Ouderaa, E., J. Schoukens and J. Renneboog (1988). "Peak factor minimisation using a time-frequency domain swapping algorithm", *IEEE Trans. Instrumentation and Measurement*, vol. 37, no. 1, pp. 145-147.

Van der Ouderaa, E. and J. Renneboog (1988). "Logtone crest factors", *IEEE Trans. Instrumentation and Measurement*, vol. 37, no. 4, pp. 656-657.

Vernet, D. (1994). "Electric cars and the quality of life", *Electro Technology*, vol. 5, no. 2, pp. 8-13.

Victor, J.D. (1979). "Nonlinear systems analysis: comparison of white noise and sum of sinusoids in a biological system", *Proc. National Academy of Science (USA)*, vol. 76, no. 2, pp. 996-998.

Victor, J.D. and B.W. Knight (1979). "Nonlinear analysis with an arbitrary stimulus ensemble", *Quarterly of Applied Mathematics*, vol. 37, no. 2, pp. 113-136.

Victor, J.D. and R. Shapley (1980). "A method of nonlinear analysis in the frequency domain", *Biophysical J.*, vol. 29, pp. 459-483.

Walter, E. and L. Pronzato (1997). *Identification of parametric models from experimental data*, Springer-Verlag, Berlin.

Weiss, M., C. Evans, D. Rees and L. Jones (1996). "Structure identification of block-oriented nonlinear systems using periodic signals", in *Proc. IEEE Instrumentation and Measurement Technology Conf.*, Belgium, vol. 1, pp. 8-13.

Wellstead, P.E. and M.B. Zarrop (1991). *Self-tuning Systems: Control and Signal Processing*, Wiley, Chichester.

Zarrop, M. (1979). *Optimal Experiment Design for Dynamic System Identification*, Lecture Notes in Control and Information Sciences No. 21, Springer-Verlag, Berlin.

Zhang, H. and S.A. Billings (1993). "Analysing non-linear systems in the frequency domain - I. The transfer function", *J. Mechanical Systems and Signal Processing*, vol. 7, no. 6, pp. 531-550.

Zhang, Q., M. Basseville and A. Benveniste (1994). "Early warning of slight changes in systems", *Automatica*, vol. 30, no. 1, pp. 95-113.